(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
14 November 2002 (14.11.2002)

PCT

(10) International Publication Number
WO 02/090512 A2

(51) International Patent Classification[7]: C12N

(21) International Application Number: PCT/US02/14538

(22) International Filing Date: 7 May 2002 (07.05.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/289,335 7 May 2001 (07.05.2001) US

(71) Applicant *(for all designated States except US)*: **GENAISSANCE PHARMACEUTICALS, INC.** [US/US]; 5 Science Park, New Haven, CT 06511 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **CHEW, Anne** [US/US]; 1477 Beacon Street #64, Brookline, MA 02446 (US). **GILSON, Christopher Raleigh** [US/US]; 30 Platt Avenue, West Haven, CT 06516 (US). **KAZEMI, Amir** [US/US]; 53 Southwind Drive, Wallingford, CT 06492 (US). **KOSHY, Beena** [IN/US]; 1298 Hartford Turnpike, Apt. 11B, North Haven, CT 06473 (US).

(74) Agents: **JUDISCH, Nathaniel D.** et al.; 5 Science Park, New Haven, CT 06511 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/090512 A2

(54) Title: HAPLOTYPES OF THE NNMT GENE

(57) Abstract: Novel genetic variants of the Nicotinamide N-Methyltransferase (NNMT) gene are described. Various genotypes, haplotypes, and haplotype pairs that exist in the general United States population are disclosed for the NNMT gene. Compositions and methods for haplotyping and/or genotyping the NNMT gene in an individual are also disclosed. Polynucleotides defined by the haplotypes disclosed herein are also described.

HAPLOTYPES OF THE NNMT GENE

RELATED APPLICATIONS

This application claims the benefit of U.S. Provisional Application Serial No. 60/289,335 filed May 7, 2001.

FIELD OF THE INVENTION

This invention relates to variation in genes that encode pharmaceutically-important proteins. In particular, this invention provides genetic variants of the human nicotinamide N-methyltransferase (NNMT) gene and methods for identifying which variant(s) of this gene is/are possessed by an individual.

BACKGROUND OF THE INVENTION

Current methods for identifying pharmaceuticals to treat disease often start by identifying, cloning, and expressing an important target protein related to the disease. A determination of whether an agonist or antagonist is needed to produce an effect that may benefit a patient with the disease is then made. Then, vast numbers of compounds are screened against the target protein to find new potential drugs. The desired outcome of this process is a lead compound that is specific for the target, thereby reducing the incidence of the undesired side effects usually caused by activity at non-intended targets. The lead compound identified in this screening process then undergoes further in vitro and in vivo testing to determine its absorption, disposition, metabolism and toxicological profiles. Typically, this testing involves use of cell lines and animal models with limited, if any, genetic diversity.

What this approach fails to consider, however, is that natural genetic variability exists between individuals in any and every population with respect to pharmaceutically-important proteins, including the protein targets of candidate drugs, the enzymes that metabolize these drugs and the proteins whose activity is modulated by such drug targets. Subtle alteration(s) in the primary nucleotide sequence of a gene encoding a pharmaceutically-important protein may be manifested as significant variation in expression, structure and/or function of the protein. Such alterations may explain the relatively high degree of uncertainty inherent in the treatment of individuals with a drug whose design is based upon a single representative example of the target or enzyme(s) involved in metabolizing the drug. For example, it is well-established that some drugs frequently have lower efficacy in some individuals than others, which means such individuals and their physicians must weigh the possible benefit of a larger dosage against a greater risk of side effects. Also, there is significant variation in how well people metabolize drugs and other exogenous chemicals, resulting in substantial interindividual variation in the toxicity and/or efficacy of such exogenous substances (Evans et al., 1999, Science 286:487-491). This variability in efficacy or toxicity of a drug in genetically-diverse patients makes many drugs ineffective or even dangerous in certain groups of the population, leading to the failure of

such drugs in clinical trials or their early withdrawal from the market even though they could be highly beneficial for other groups in the population. This problem significantly increases the time and cost of drug discovery and development, which is a matter of great public concern.

It is well-recognized by pharmaceutical scientists that considering the impact of the genetic variability of pharmaceutically-important proteins in the early phases of drug discovery and development is likely to reduce the failure rate of candidate and approved drugs (Marshall A 1997 Nature Biotech 15:1249-52; Kleyn PW et al. 1998 Science 281: 1820-21; Kola I 1999 Curr Opin Biotech 10:589-92; Hill AVS et al. 1999 in Evolution in Health and Disease Stearns SS (Ed.) Oxford University Press, New York, pp 62-76; Meyer U.A. 1999 in Evolution in Health and Disease Stearns SS (Ed.) Oxford University Press, New York, pp 41-49; Kalow W et al. 1999 Clin. Pharm. Therap. 66:445-7; Marshall, E 1999 Science 284:406-7; Judson R et al. 2000 Pharmacogenomics 1:1-12; Roses AD 2000 Nature 405:857-65). However, in practice this has been difficult to do, in large part because of the time and cost required for discovering the amount of genetic variation that exists in the population (Chakravarti A 1998 Nature Genet 19:216-7; Wang DG et al 1998 Science 280:1077-82; Chakravarti A 1999 Nat Genet 21:56-60 (suppl); Stephens JC 1999 Mol. Diagnosis 4:309-317; Kwok PY and Gu S 1999 Mol. Med. Today 5:538-43; Davidson S 2000 Nature Biotech 18:1134-5).

The standard for measuring genetic variation among individuals is the haplotype, which is the ordered combination of polymorphisms in the sequence of each form of a gene that exists in the population. Because haplotypes represent the variation across each form of a gene, they provide a more accurate and reliable measurement of genetic variation than individual polymorphisms. For example, while specific variations in gene sequences have been associated with a particular phenotype such as disease susceptibility (Roses AD supra; Ulbrecht M et al. 2000 Am J Respir Crit Care Med 161: 469-74) and drug response (Wolfe CR et al. 2000 BMJ 320:987-90; Dahl BS 1997 Acta Psychiatr Scand 96 (Suppl 391): 14-21), in many other cases an individual polymorphism may be found in a variety of genomic backgrounds, i.e., different haplotypes, and therefore shows no definitive coupling between the polymorphism and the causative site for the phenotype (Clark AG et al. 1998 Am J Hum Genet 63:595-612; Ulbrecht M et al. 2000 supra; Drysdale et al. 2000 PNAS 97:10483-10488). Thus, there is an unmet need in the pharmaceutical industry for information on what haplotypes exist in the population for pharmaceutically-important genes. Such haplotype information would be useful in improving the efficiency and output of several steps in the drug discovery and development process, including target validation, identifying lead compounds, and early phase clinical trials (Marshall et al., supra).

One pharmaceutically-important gene for the treatment of Parkinson's disease and cancer cachexia is the nicotinamide N-methyltransferase (NNMT) gene or its encoded product. NNMT catalyzes the N-methylation of nicotinamide and other pyridines to form pyridinium ions (SWISS-PROT: P40261). This activity is important for the biotransformation of drugs and xenobiotic compounds. The phenotypic variability of NNMT activity prompted the hypothesis that the enzyme

may be regulated by polymorphisms (Aksoy et al., *J Biol Chem.* 1994; 269:14835-14840). Differences in NNMT activity may lead to variable N-methylation of pyridine compounds and to individual differences in toxicity (Rini et al., *Clin Chim. Acta* 1990; 186:359-374).

For example, a person with a higher level of NNMT activity may be predisposed to the neurodegenerative disorder Parkinson's disease (PD) (Parsons et al., *J Neuropathol. Exp Neurol.* 2002; 61:111-124; Aoyama et al., *Neurosci Lett.* 2001; 298:78-80). PD patients showed an increase in NNMT protein levels in both their cerebrospinal fluid and in those brain regions relevant to PD pathology. Furthermore, NNMT activity measured in relevant brain regions was increased in PD patients as compared to non-PD controls. NNMT is also thought to be associated with the neurons that degenerate in PD, as NNMT expression decreases with disease duration. Taken together, these data suggest that NNMT is a candidate gene for PD.

NNMT may also be a marker for cancer cachexia, a common cause of death in advanced cancer patients (Okamura et al., *Jpn. J Cancer Res.* 1998; 89:649-656; Barber et al., *Surg. Oncol.* 1999; 8:133-141). While cachexia is a general state of physical ill health associated with chronic disese, cachexia associated with cancer is specifically characterized by metabolic abnormalities that can result in severe weight loss. Mice in which cachexia was induced showed a marked progessive increase in liver NNMT activity that paralleled weight loss, and continued until death. Agents that inhibited NNMT activity in these mice also prevented weight loss. Therefore, therapeutics that target NNMT may be useful in treating cancer cachexia.

The nicotinamide N-methyltransferase gene is located on chromosome 11q23.1 and contains 3 exons that encode a 264 amino acid protein. A reference sequence for the NNMT gene comprises the non-contigous sequences shown in the contiguous lines of Figure 1, which is a composite genomic sequence based on Genaissance Reference No. 447335 (SEQ ID NO: 1). Reference sequences for the coding sequence (GenBank Accession No. NM_006169.1) and protein are shown in Figures 2 (SEQ ID NO: 2) and 3 (SEQ ID NO: 3), respectively.

Because of the potential for variation in the NNMT gene to affect the expression and function of the encoded protein, it would be useful to know whether polymorphisms exist in the NNMT gene, as well as how such polymorphisms are combined in different copies of the gene. Such information could be applied for studying the biological function of NNMT as well as in identifying drugs metabolized by this protein or drugs targeting this protein for the treatment of disorders related to its abnormal expression or function.

SUMMARY OF THE INVENTION

Accordingly, the inventors herein have discovered 3 novel polymorphic sites in the NNMT gene. These polymorphic sites (PS) correspond to the following nucleotide positions in Figure 1: 394 (PS1), 928 (PS2) and 2696 (PS3). The polymorphisms at these sites are adenine or thymine at PS1, thymine or cytosine at PS2 and thymine or cytosine at PS3. In addition, the inventors have determined

the identity of the alleles at these sites in a human reference population of 79 unrelated individuals self-identified as belonging to one of four major population groups: African descent, Asian, Caucasian and Hispanic/Latino. From this information, the inventors deduced a set of haplotypes and haplotype pairs for PS1-PS3 in the NNMT gene, which are shown below in Tables 4 and 3, respectively. Each of these NNMT haplotypes constitutes a code, or genetic marker, that defines the variant nucleotides that exist in the human population at this set of polymorphic sites in the NNMT gene. Thus each NNMT haplotype also represents a naturally-occurring isoform (also referred to herein as an "isogene") of the NNMT gene. The frequency of each haplotype and haplotype pair within the total reference population and within each of the four major population groups included in the reference population was also determined.

Thus, in one embodiment, the invention provides a method, composition and kit for genotyping the NNMT gene in an individual. The genotyping method comprises identifying the nucleotide pair that is present at one or more polymorphic sites selected from the group consisting of PS1, PS2 and PS3 in both copies of the NNMT gene from the individual. A genotyping composition of the invention comprises an oligonucleotide probe or primer which is designed to specifically hybridize to a target region containing, or adjacent to, one of these NNMT polymorphic sites. In one embodiment, a genotyping kit of the invention comprises a set of oligonucleotides designed to genotype each of these novel NNMT polymorphic sites. The genotyping method, composition, and kit are useful in determining whether an individual has one of the haplotypes in Table 4 below or has one of the haplotype pairs in Table 3 below.

The invention also provides a method for haplotyping the NNMT gene in an individual. In one embodiment, the haplotyping method comprises determining, for one copy of the NNMT gene, the identity of the nucleotide at one or more polymorphic sites selected from the group consisting of PS1, PS2 and PS3. In another embodiment, the haplotyping method comprises determining whether one copy of the individual's NNMT gene is defined by one of the NNMT haplotypes shown in Table 4, below, or a sub-haplotype thereof. In a preferred embodiment, the haplotyping method comprises determining whether both copies of the individual's NNMT gene are defined by one of the NNMT haplotype pairs shown in Table 3 below, or a sub-haplotype pair thereof. Establishing the NNMT haplotype or haplotype pair of an individual is useful for improving the efficiency and reliability of several steps in the discovery and development of drugs metabolized by NNMT or drugs for treating diseases associated with NNMT activity, e.g., Parkinson's disease and cancer cachexia.

For example, the haplotyping method can be used by the pharmaceutical research scientist to validate NNMT as a candidate target for treating a specific condition or disease predicted to be associated with NNMT activity. Determining for a particular population the frequency of one or more of the individual NNMT haplotypes or haplotype pairs described herein will facilitate a decision on whether to pursue NNMT as a target for treating the specific disease of interest. In particular, if variable NNMT activity is associated with the disease, then one or more NNMT haplotypes or

haplotype pairs will be found at a higher frequency in disease cohorts than in appropriately genetically matched controls. Conversely, if each of the observed NNMT haplotypes are of similar frequencies in the disease and control groups, then it may be inferred that variable NNMT activity has little, if any, involvement with that disease. In either case, the pharmaceutical research scientist can, without *a priori* knowledge as to the phenotypic effect of any NNMT haplotype or haplotype pair, apply the information derived from detecting NNMT haplotypes in an individual to decide whether modulating NNMT activity would be useful in treating the disease.

The claimed invention is also useful in screening for compounds targeting NNMT to treat a specific condition or disease predicted to be associated with NNMT activity. For example, detecting which of the NNMT haplotypes or haplotype pairs disclosed herein are present in individual members of a population with the specific disease of interest enables the pharmaceutical scientist to screen for a compound(s) that displays the highest desired agonist or antagonist activity for each of the NNMT isoforms present in the disease population, or for only the most frequent NNMT isoforms present in the disease population. Thus, without requiring any *a priori* knowledge of the phenotypic effect of any particular NNMT haplotype or haplotype pair, the claimed haplotyping method provides the scientist with a tool to identify lead compounds that are more likely to show efficacy in clinical trials.

Haplotyping the NNMT gene in an individual is also useful to control for genetically-based bias in the design of candidate drugs that target or are metabolized by NNMT. For example, for a lead compound that is metabolized by NNMT, the pharmaceutical scientist of ordinary skill would be concerned that a favorable efficacy and/or side effect profile shown in a Phase II or III trial may not be replicated in the general population if a higher (or lower) percentage of patients in the treatment group, compared to the general population, have a form of the NNMT gene that makes them genetically predisposed to metabolize the drug more efficiently than patients with other forms of the NNMT gene. Similarly, this pharmaceutical scientist would recognize the potential for bias in the results of a Phase II or Phase III clinical trial of a drug targeting NNMT that could be introduced if individuals whose NNMT gene structure makes them genetically predisposed to respond well to the drug are present in a higher (or lower) frequency in the treatment group than in the control group (Bacanu et al., 2000, *AM. J. Hum. Gen.* **66**:1933-44; Pritchard et al., 2000, *Am. J. Hum. Gen.* **67**: 170-81).

The pharmaceutical scientist can immediately reduce this potential for genetically-based bias in the results of clinical trials of drugs metabolized by or targeting NMT by practicing the claimed invention. In particular, by determining which of the NNMT haplotypes disclosed herein are present in individuals recruited to participate in a clinical trial of a drug metabolized by or targeting NNMT, the pharmaceutical scientist can then assign that individual to the treatment or control group as appropriate to ensure that approximately equal frequencies of different NNMT haplotypes (or haplotype pairs) are represented in the two groups and/or the frequencies of different NNMT haplotypes or haplotype pairs are similar to the frequencies in the general population. Thus, by practicing the claimed invention, the pharmaceutical scientist can more confidently rely on the

information learned from the trial, without first determining the phenotypic effect of any NNMT haplotype or haplotype pair.

In another embodiment, the invention provides a method for identifying an association between a trait and a NNMT genotype, haplotype, or haplotype pair for one or more of the novel polymorphic sites described herein. The method comprises comparing the frequency of the NNMT genotype, haplotype, or haplotype pair in a population exhibiting the trait with the frequency of the NNMT genotype or haplotype in a reference population. A different frequency of the NNMT genotype, haplotype, or haplotype pair in the trait population than in the reference population indicates the trait is associated with the NNMT genotype, haplotype, or haplotype pair. In preferred embodiments, the trait is susceptibility to a disease, severity of a disease, the staging of a disease or response to a drug. In a particularly preferred embodiment, the NNMT haplotype is selected from the haplotypes shown in Table 4, or a sub-haplotype thereof. Such methods have applicability in developing diagnostic tests for assessing potential drug metabolism by NNMT and for developing diagnostic tests and therapeutic treatments for Parkinson's disease or cancer cachexia.

In yet another embodiment, the invention provides an isolated polynucleotide comprising a nucleotide sequence which is a polymorphic variant of a reference sequence for the NNMT gene or a fragment thereof. The reference sequence comprises the contiguous sequences shown in Figure 1 and the polymorphic variant comprises at least one polymorphism selected from the group consisting of thymine at PS1, cytosine at PS2 and cytosine at PS3.

A particularly preferred polymorphic variant is an isogene of the NNMT gene. A NNMT isogene of the invention comprises adenine or thymine at PS1, thymine or cytosine at PS2 and thymine or cytosine at PS3. The invention also provides a collection of NNMT isogenes, referred to herein as a NNMT genome anthology.

In another embodiment, the invention provides a polynucleotide comprising a polymorphic variant of a reference sequence for a NNMT cDNA or a fragment thereof. The reference sequence comprises SEQ ID NO:2 (Fig.2) and the polymorphic cDNA comprises cytosine at a position corresponding to nucleotide 426. A particularly preferred polymorphic cDNA variant is A represented in Table 7.

Polynucleotides complementary to these NNMT genomic and cDNA variants are also provided by the invention. It is believed that polymorphic variants of the NNMT gene will be useful in studying the expression and function of NNMT, and in expressing NNMT protein for use in screening for candidate drugs that may be metabolized by NNMT or to treat diseases related to NNMT activity.

In other embodiments, the invention provides a recombinant expression vector comprising one of the polymorphic genomic and cDNA variants operably linked to expression regulatory elements as well as a recombinant host cell transformed or transfected with the expression vector. The recombinant vector and host cell may be used to express NNMT for protein structure analysis and

drug binding studies.

The present invention also provides nonhuman transgenic animals comprising one or more of the NNMT polymorphic genomic variants described herein and methods for producing such animals. The transgenic animals are useful for studying expression of the NNMT isogenes *in vivo*, for *in vivo* screening and testing of drugs targeted against NNMT protein, and for testing the efficacy of therapeutic agents and compounds for Parkinson's disease and cancer cachexia in a biological system.

The present invention also provides a computer system for storing and displaying polymorphism data determined for the NNMT gene. The computer system comprises a computer processing unit; a display; and a database containing the polymorphism data. The polymorphism data includes one or more of the following: the polymorphisms, the genotypes, the haplotypes, and the haplotype pairs identified for the NNMT gene in a reference population. In a preferred embodiment, the computer system is capable of producing a display showing NNMT haplotypes organized according to their evolutionary relationships.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 illustrates a reference sequence for the NNMT gene (Genaissance Reference No. 447335; contiguous lines), with the start and stop positions of each region of coding sequence indicated with a bracket ([or]) and the numerical position below the sequence and the polymorphic site(s) and polymorphism(s) identified by Applicants in a reference population indicated by the variant nucleotide positioned below the polymorphic site in the sequence. SEQ ID NO:1 is equivalent to Figure 1, with the two alternative allelic variants of each polymorphic site indicated by the appropriate nucleotide symbol (R= G or A, Y= T or C, M= A or C, K= G or T, S= G or C, and W= A or T; WIPO standard ST.25). SEQ ID NO:21 is a modified version of SEQ ID NO:1 that shows the context sequence of each polymorphic site, PS1-PS3, in a uniform format to facilitate electronic searching. For each polymorphic site, SEQ ID NO:21 contains a block of 60 bases of the nucleotide sequence encompassing the centrally-located polymorphic site at the 30th position, followed by 60 bases of unspecified sequence to represent that each PS is separated by genomic sequence whose composition is defined elsewhere herein.

Figure 2 illustrates a reference sequence for the NNMT coding sequence (contiguous lines; SEQ ID NO:2), with the polymorphic site(s) and polymorphism(s) identified by Applicants in a reference population indicated by the variant nucleotide positioned below the polymorphic site in the sequence.

Figure 3 illustrates a reference sequence for the NNMT protein (contiguous lines; SEQ ID NO:3).

DESCRIPTION OF THE PREFERRED EMBODIMENTS

The present invention is based on the discovery of novel variants of the NNMT gene. As described in more detail below, the inventors herein discovered 5 isogenes of the NNMT gene by

characterizing the NNMT gene found in genomic DNAs isolated from an Index Repository that contains immortalized cell lines from one chimpanzee and 93 human individuals. The human individuals included a reference population of 79 unrelated individuals self-identified as belonging to one of four major population groups: Caucasian (21 individuals), African descent (20 individuals), Asian (20 individuals), or Hispanic/Latino (18 individuals). To the extent possible, the members of this reference population were organized into population subgroups by their self-identified ethnogeographic origin as shown in Table 1 below. In addition, the Index Repository contains three unrelated indigenous American Indians (one from each of North, Central and South America), one three-generation Caucasian family (from the CEPH Utah cohort) and one two-generation African-American family.

Table 1. Population Groups in the Index Repository

Population Group	Population Subgroup	No. of Individuals
African descent		20
	Sierra Leone	1
Asian		20
	Burma	1
	China	3
	Japan	6
	Korea	1
	Philippines	5
	Vietnam	4
Caucasian		21
	British Isles	3
	British Isles/Central	4
	British Isles/Eastern	1
	Central/Eastern	1
	Eastern	3
	Central/Mediterranean	1
	Mediterranean	2
	Scandinavian	2
Hispanic/Latino		18
	Caribbean	8
	Caribbean (Spanish Descent)	2
	Central American (Spanish Descent)	1
	Mexican American	4
	South American (Spanish Descent)	3

The NNMT isogenes present in the human reference population are defined by haplotypes for 3 polymorphic sites in the NNMT gene, all of which are believed to be novel. The novel NNMT polymorphic sites identified by the inventors are referred to as PS1-PS3 to designate the order in which they are located in the gene (see Table 2 below). Using the genotypes identified in the Index Repository for PS1-PS3 and the methodology described in the Examples below, the inventors herein also determined the pair of haplotypes for the NNMT gene present in individual human members of this repository. The human genotypes and haplotypes found in the repository for the NNMT gene

include those shown in Tables 3 and 4, respectively. The polymorphism and haplotype data disclosed herein are useful for validating whether NNMT is a suitable target for drugs to treat Parkinson's disease and cancer cachexia, screening for such drugs and reducing bias in clinical trials of such drugs. These data are also useful to control for genetically-based bias in the design of drugs metabolized by NNMT.

In the context of this disclosure, the following terms shall be defined as follows unless otherwise indicated:

Allele - A particular form of a genetic locus, distinguished from other forms by its particular nucleotide sequence.

Candidate Gene – A gene which is hypothesized to be responsible for a disease, condition, or the response to a treatment, or to be correlated with one of these.

Gene - A segment of DNA that contains the coding sequence for a protein, wherein the segment may include promoters, exons, introns, and other untranslated regions that control expression.

Genotype – An unphased 5′ to 3′ sequence of nucleotide pair(s) found at one or more polymorphic sites in a locus on a pair of homologous chromosomes in an individual. As used herein, genotype includes a full-genotype and/or a sub-genotype as described below.

Full-genotype – The unphased 5′ to 3′ sequence of nucleotide pairs found at all polymorphic sites examined herein in a locus on a pair of homologous chromosomes in a single individual.

Sub-genotype – The unphased 5′ to 3′ sequence of nucleotides seen at a subset of the polymorphic sites examined herein in a locus on a pair of homologous chromosomes in a single individual.

Genotyping – A process for determining a genotype of an individual.

Haplotype – A 5′ to 3′ sequence of nucleotides found at one or more polymorphic sites in a locus on a single chromosome from a single individual. As used herein, haplotype includes a full-haplotype and/or a sub-haplotype as described below.

Full-haplotype – The 5′ to 3′ sequence of nucleotides found at all polymorphic sites examined herein in a locus on a single chromosome from a single individual.

Sub-haplotype – The 5′ to 3′ sequence of nucleotides seen at a subset of the polymorphic sites examined herein in a locus on a single chromosome from a single individual.

Haplotype pair – The two haplotypes found for a locus in a single individual.

Haplotyping – A process for determining one or more haplotypes in an individual and includes use of family pedigrees, molecular techniques and/or statistical inference.

Haplotype data - Information concerning one or more of the following for a specific gene: a listing of the haplotype pairs in each individual in a population; a listing of the different haplotypes in a population; frequency of each haplotype in that or other populations, and any known associations between one or more haplotypes and a trait.

Isoform – A particular form of a gene, mRNA, cDNA, coding sequence or the protein

encoded thereby, distinguished from other forms by its particular sequence and/or structure.

Isogene – One of the isoforms (e.g., alleles) of a gene found in a population. An isogene (or allele) contains all of the polymorphisms present in the particular isoform of the gene.

Isolated – As applied to a biological molecule such as RNA, DNA, oligonucleotide, or protein, isolated means the molecule is substantially free of other biological molecules such as nucleic acids, proteins, lipids, carbohydrates, or other material such as cellular debris and growth media. Generally, the term "isolated" is not intended to refer to a complete absence of such material or to absence of water, buffers, or salts, unless they are present in amounts that substantially interfere with the methods of the present invention.

Locus - A location on a chromosome or DNA molecule corresponding to a gene or a physical or phenotypic feature, where physical features include polymorphic sites.

Naturally-occurring – A term used to designate that the object it is applied to, e.g., naturally-occurring polynucleotide or polypeptide, can be isolated from a source in nature and which has not been intentionally modified by man.

Nucleotide pair – The nucleotides found at a polymorphic site on the two copies of a chromosome from an individual.

Phased – As applied to a sequence of nucleotide pairs for two or more polymorphic sites in a locus, phased means the combination of nucleotides present at those polymorphic sites on a single copy of the locus is known.

Polymorphic site (PS) – A position on a chromosome or DNA molecule at which at least two alternative sequences are found in a population.

Polymorphic variant (variant)– A gene, mRNA, cDNA, polypeptide, protein or peptide whose nucleotide or amino acid sequence varies from a reference sequence due to the presence of a polymorphism in the gene.

Polymorphism – The sequence variation observed in an individual at a polymorphic site. Polymorphisms include nucleotide substitutions, insertions, deletions and microsatellites and may, but need not, result in detectable differences in gene expression or protein function.

Polymorphism data – Information concerning one or more of the following for a specific gene: location of polymorphic sites; sequence variation at those sites; frequency of polymorphisms in one or more populations; the different genotypes and/or haplotypes determined for the gene; frequency of one or more of these genotypes and/or haplotypes in one or more populations; any known association(s) between a trait and a genotype or a haplotype for the gene.

Polymorphism Database – A collection of polymorphism data arranged in a systematic or methodical way and capable of being individually accessed by electronic or other means.

Polynucleotide – A nucleic acid molecule comprised of single-stranded RNA or DNA or comprised of complementary, double-stranded DNA.

Population Group – A group of individuals sharing a common ethnogeographic origin.

Reference Population – A group of subjects or individuals who are predicted to be representative of the genetic variation found in the general population. Typically, the reference population represents the genetic variation in the population at a certainty level of at least 85%, preferably at least 90%, more preferably at least 95% and even more preferably at least 99%.

Single Nucleotide Polymorphism (SNP) – Typically, the specific pair of nucleotides observed at a single polymorphic site. In rare cases, three or four nucleotides may be found.

Subject – A human individual whose genotypes or haplotypes or response to treatment or disease state are to be determined.

Treatment - A stimulus administered internally or externally to a subject.

Unphased – As applied to a sequence of nucleotide pairs for two or more polymorphic sites in a locus, unphased means the combination of nucleotides present at those polymorphic sites on a single copy of the locus is not known.

As discussed above, information on the identity of genotypes and haplotypes for the NNMT gene of any particular individual as well as the frequency of such genotypes and haplotypes in any particular population of individuals is useful for a variety of drug discovery and development applications. Thus, the invention also provides compositions and methods for detecting the novel NNMT polymorphisms, haplotypes and haplotype pairs identified herein.

The compositions comprise at least one oligonucleotide for detecting the variant nucleotide or nucleotide pair located at a NNMT polymorphic site in one copy or two copies of the NNMT gene. Such oligonucleotides are referred to herein as NNMT haplotyping oligonucleotides or genotyping oligonucleotides, respectively, and collectively as NNMT oligonucleotides. In one embodiment, a NNMT haplotyping or genotyping oligonucleotide is a probe or primer capable of hybridizing to a target region that contains, or that is located close to, one of the novel polymorphic sites described herein.

As used herein, the term "oligonucleotide" refers to a polynucleotide molecule having less than about 100 nucleotides. A preferred oligonucleotide of the invention is 10 to 35 nucleotides long. More preferably, the oligonucleotide is between 15 and 30, and most preferably, between 20 and 25 nucleotides in length. The exact length of the oligonucleotide will depend on many factors that are routinely considered and practiced by the skilled artisan. The oligonucleotide may be comprised of any phosphorylation state of ribonucleotides, deoxyribonucleotides, and acyclic nucleotide derivatives, and other functionally equivalent derivatives. Alternatively, oligonucleotides may have a phosphate-free backbone, which may be comprised of linkages such as carboxymethyl, acetamidate, carbamate, polyamide (peptide nucleic acid (PNA)) and the like (Varma, R. in Molecular Biology and Biotechnology, A Comprehensive Desk Reference, Ed. R. Meyers, VCH Publishers, Inc. (1995), pages 617-620). Oligonucleotides of the invention may be prepared by chemical synthesis using any suitable methodology known in the art, or may be derived from a biological sample, for example, by restriction digestion. The oligonucleotides may be labeled, according to any technique known in the

art, including use of radiolabels, fluorescent labels, enzymatic labels, proteins, haptens, antibodies, sequence tags and the like.

Haplotyping or genotyping oligonucleotides of the invention must be capable of specifically hybridizing to a target region of a NNMT polynucleotide. Preferably, the target region is located in a NNMT isogene. As used herein, specific hybridization means the oligonucleotide forms an anti-parallel double-stranded structure with the target region under certain hybridizing conditions, while failing to form such a structure when incubated with another region in the NNMT polynucleotide or with a non-NNMT polynucleotide under the same hybridizing conditions. Preferably, the oligonucleotide specifically hybridizes to the target region under conventional high stringency conditions. The skilled artisan can readily design and test oligonucleotide probes and primers suitable for detecting polymorphisms in the NNMT gene using the polymorphism information provided herein in conjunction with the known sequence information for the NNMT gene and routine techniques.

A nucleic acid molecule such as an oligonucleotide or polynucleotide is said to be a "perfect" or "complete" complement of another nucleic acid molecule if every nucleotide of one of the molecules is complementary to the nucleotide at the corresponding position of the other molecule. A nucleic acid molecule is "substantially complementary" to another molecule if it hybridizes to that molecule with sufficient stability to remain in a duplex form under conventional low-stringency conditions. Conventional hybridization conditions are described, for example, by Sambrook J. et al., in Molecular Cloning, A Laboratory Manual, 2nd Edition, Cold Spring Harbor Press, Cold Spring Harbor, NY (1989) and by Haymes, B.D. et al. in Nucleic Acid Hybridization, A Practical Approach, IRL Press, Washington, D.C. (1985). While perfectly complementary oligonucleotides are preferred for detecting polymorphisms, departures from complete complementarity are contemplated where such departures do not prevent the molecule from specifically hybridizing to the target region. For example, an oligonucleotide primer may have a non-complementary fragment at its 5′ end, with the remainder of the primer being complementary to the target region. Alternatively, non-complementary nucleotides may be interspersed into the probe or primer as long as the resulting probe or primer is still capable of specifically hybridizing to the target region.

Preferred haplotyping or genotyping oligonucleotides of the invention are allele-specific oligonucleotides. As used herein, the term allele-specific oligonucleotide (ASO) means an oligonucleotide that is able, under sufficiently stringent conditions, to hybridize specifically to one allele of a gene, or other locus, at a target region containing a polymorphic site while not hybridizing to the corresponding region in another allele(s). As understood by the skilled artisan, allele-specificity will depend upon a variety of readily optimized stringency conditions, including salt and formamide concentrations, as well as temperatures for both the hybridization and washing steps. Examples of hybridization and washing conditions typically used for ASO probes are found in Kogan et al., "Genetic Prediction of Hemophilia A" in PCR Protocols, A Guide to Methods and Applications, Academic Press, 1990 and Ruaño et al., 87 *Proc. Natl. Acad. Sci. USA* 6296-6300, 1990. Typically, an

ASO will be perfectly complementary to one allele while containing a single mismatch for another allele.

Allele-specific oligonucleotides of the invention include ASO probes and ASO primers. ASO probes which usually provide good discrimination between different alleles are those in which a central position of the oligonucleotide probe aligns with the polymorphic site in the target region (e.g., approximately the 7th or 8th position in a 15mer, the 8th or 9th position in a 16mer, and the 10th or 11th position in a 20mer). An ASO primer of the invention has a 3′ terminal nucleotide, or preferably a 3′ penultimate nucleotide, that is complementary to only one nucleotide of a particular SNP, thereby acting as a primer for polymerase-mediated extension only if the allele containing that nucleotide is present. ASO probes and primers hybridizing to either the coding or noncoding strand are contemplated by the invention. ASO probes and primers listed below use the appropriate nucleotide symbol (R= G or A, Y= T or C, M= A or C, K= G or T, S= G or C, and W= A or T; WIPO standard ST.25) at the position of the polymorphic site to represent that the ASO contains either of the two alternative allelic variants observed at that polymorphic site.

A preferred ASO probe for detecting NNMT gene polymorphisms comprises a nucleotide sequence, listed 5′ to 3′, selected from the group consisting of:

CGAGCTCWAGTGCTC (SEQ ID NO:4) and its complement,
AGTCATAYAGATGGA (SEQ ID NO:5) and its complement, and
AGTGTGAYGTGACTC (SEQ ID NO:6) and its complement.

A preferred ASO primer for detecting NNMT gene polymorphisms comprises a nucleotide sequence, listed 5′ to 3′, selected from the group consisting of:

TCCTGACGAGCTCWA (SEQ ID NO:7); CAGAGGGAGCACTWG (SEQ ID NO:8);
AATGTGAGTCATAYA (SEQ ID NO:9); TGAGACTCCATCTRT (SEQ ID NO:10);
TGCTGAAGTGTGAYG (SEQ ID NO:11) and GGCTCTGAGTCACRT (SEQ ID NO:12).

Other oligonucleotides of the invention hybridize to a target region located one to several nucleotides downstream of one of the novel polymorphic sites identified herein. Such oligonucleotides are useful in polymerase-mediated primer extension methods for detecting one of the novel polymorphisms described herein and therefore such oligonucleotides are referred to herein as "primer-extension oligonucleotides". In a preferred embodiment, the 3′-terminus of a primer-extension oligonucleotide is a deoxynucleotide complementary to the nucleotide located immediately adjacent to the polymorphic site.

A particularly preferred oligonucleotide primer for detecting NNMT gene polymorphisms by primer extension terminates in a nucleotide sequence, listed 5′ to 3′, selected from the group consisting of:

TGACGAGCTC (SEQ ID NO:13); AGGGAGCACT (SEQ ID NO:14);
GTGAGTCATA (SEQ ID NO:15); GACTCCATCT (SEQ ID NO:16);

TGAAGTGTGA (SEQ ID NO:17) and TCTGAGTCAC (SEQ ID NO:18).

In some embodiments, a composition contains two or more differently labeled NNMT oligonucleotides for simultaneously probing the identity of nucleotides or nucleotide pairs at two or more polymorphic sites. It is also contemplated that primer compositions may contain two or more sets of allele-specific primer pairs to allow simultaneous targeting and amplification of two or more regions containing a polymorphic site.

NNMT oligonucleotides of the invention may also be immobilized on or synthesized on a solid surface such as a microchip, bead, or glass slide (see, e.g., WO 98/20020 and WO 98/20019). Such immobilized oligonucleotides may be used in a variety of polymorphism detection assays, including but not limited to probe hybridization and polymerase extension assays. Immobilized NNMT oligonucleotides of the invention may comprise an ordered array of oligonucleotides designed to rapidly screen a DNA sample for polymorphisms in multiple genes at the same time.

In another embodiment, the invention provides a kit comprising at least two NNMT oligonucleotides packaged in separate containers. The kit may also contain other components such as hybridization buffer (where the oligonucleotides are to be used as a probe) packaged in a separate container. Alternatively, where the oligonucleotides are to be used to amplify a target region, the kit may contain, packaged in separate containers, a polymerase and a reaction buffer optimized for primer extension mediated by the polymerase, such as PCR.

The above described oligonucleotide compositions and kits are useful in methods for genotyping and/or haplotyping the NNMT gene in an individual. As used herein, the terms "NNMT genotype" and "NNMT haplotype" mean the genotype or haplotype contains the nucleotide pair or nucleotide, respectively, that is present at one or more of the novel polymorphic sites described herein and may optionally also include the nucleotide pair or nucleotide present at one or more additional polymorphic sites in the NNMT gene. The additional polymorphic sites may be currently known polymorphic sites or sites that are subsequently discovered.

One embodiment of a genotyping method of the invention involves examining both copies of the individual's NNMT gene, or a fragment thereof, to identify the nucleotide pair at one or more polymorphic sites selected from the group consisting of PS1, PS2 and PS3 in the two copies to assign a NNMT genotype to the individual. In some embodiments, "examining a gene" may include examining one or more of: DNA containing the gene, mRNA transcripts thereof, or cDNA copies thereof. As will be readily understood by the skilled artisan, the two "copies" of a gene, mRNA or cDNA (or fragment of such NNMT molecules) in an individual may be the same allele or may be different alleles. In another embodiment, a genotyping method of the invention comprises determining the identity of the nucleotide pair at each of PS1-PS3.

One method of examining both copies of the individual's NNMT gene is by isolating from the individual a nucleic acid sample comprising the two copies of the NNMT gene, mRNA transcripts

thereof or cDNA copies thereof, or a fragment of any of the foregoing, that are present in the individual. Typically, the nucleic acid sample is isolated from a biological sample taken from the individual, such as a blood sample or tissue sample. Suitable tissue samples include whole blood, semen, saliva, tears, urine, fecal material, sweat, buccal, skin and hair. The nucleic acid sample may be comprised of genomic DNA, mRNA, or cDNA and, in the latter two cases, the biological sample must be obtained from a tissue in which the NNMT gene is expressed. Furthermore it will be understood by the skilled artisan that mRNA or cDNA preparations would not be used to detect polymorphisms located in introns or in 5′ and 3′ untranslated regions if not present in the mRNA or cDNA. If a NNMT gene fragment is isolated, it must contain the polymorphic site(s) to be genotyped.

One embodiment of a haplotyping method of the invention comprises examining one copy of the individual's NNMT gene, or a fragment thereof, to identify the nucleotide at one or more polymorphic sites selected from the group consisting of PS1, PS2 and PS3 in that copy to assign a NNMT haplotype to the individual. In a preferred embodiment, the nucleotide at each of PS1-PS3 is identified. In a particularly preferred embodiment, the NNMT haplotype assigned to the individual is selected from the group consisting of the NNMT haplotypes shown in Table 4.

In some embodiments, "examining a gene" may include examining one or more of: DNA containing the gene, mRNA transcripts thereof, or cDNA copies thereof. One method of examining one copy of the individual's NNMT gene is by isolating from the individual a nucleic acid sample containing only one of the two copies of the NNMT gene, mRNA or cDNA, or a fragment of such NNMT molecules, that is present in the individual and determining in that copy the identity of the nucleotide at one or more polymorphic sites selected from the group consisting of PS1, PS2 and PS3 in that copy to assign a NNMT haplotype to the individual. In a particularly preferred embodiment, the nucleotide at each of PS1-PS3 is identified.

The nucleic acid used in the above haplotyping methods of the invention may be isolated using any method capable of separating the two copies of the NNMT gene or fragment such as one of the methods described above for preparing NNMT isogenes, with targeted *in vivo* cloning being the preferred approach. As will be readily appreciated by those skilled in the art, any individual clone will typically only provide haplotype information on one of the two NNMT gene copies present in an individual. If haplotype information is desired for the individual's other copy, additional NNMT clones will usually need to be examined. Typically, at least five clones should be examined to have more than a 90% probability of haplotyping both copies of the NNMT gene in an individual. In some cases, however, once the haplotype for one NNMT allele is directly determined, the haplotype for the other allele may be inferred if the individual has a known genotype for the polymorphic sites of interest or if the haplotype frequency or haplotype pair frequency for the individual's population group is known.

In another embodiment, the haplotyping method comprises determining whether an individual has one or more of the NNMT haplotypes shown in Table 4. This can be accomplished by identifying

the phased sequence of nucleotides present at PS1-PS3 for at least one copy of the individual's NNMT gene and assigning to that copy a NNMT haplotype that is consistent with the phased sequence, wherein the NNMT haplotype is selected from the group consisting of the NNMT haplotypes shown in Table 4 and wherein each of the NNMT haplotypes in Table 4 comprises a sequence of polymorphisms whose positions and identities are set forth in the table. This identifying step does not necessarily require that each of PS1-PS3 be directly examined. Typically only a subset of PS1-PS3 will need to be directly examined to assign to an individual one or more of the haplotypes shown in Table 4. This is because at least one polymorphic site in a gene is frequently in strong linkage disequilibrium with one or more other polymorphic sites in that gene (Drysdale, CM et al. 2000 *PNAS* 97:10483-10488; Rieder MJ et al. 1999 *Nature Genetics* 22:59-62). Two nucleotide alleles are said to be in linkage disequilibrium if the presence of a particular allele at one polymorphic site predicts the presence of the other allele at a second polymorphic site (Stephens, JC 1999, *Mol. Diag.* 4:309-317). Techniques for determining whether any two polymorphic sites are in linkage disequilibrium are well-known in the art (Weir B.S. 1996 *Genetic Data Analysis II*, Sinauer Associates, Inc. Publishers, Sunderland, MA). In addition, Johnson et al. (2001 *Nature Genetics* 29: 233-237) presented one possible method for selection of subsets of polymorphic sites suitable for identifying known haplotypes.

In another embodiment of a haplotyping method of the invention, a NNMT haplotype pair is determined for an individual by identifying the phased sequence of nucleotides at one or more polymorphic sites selected from the group consisting of PS1, PS2 and PS3 in each copy of the NNMT gene that is present in the individual. In a particularly preferred embodiment, the haplotyping method comprises identifying the phased sequence of nucleotides at each of PS1-PS3 in each copy of the NNMT gene.

In another embodiment, the haplotyping method comprises determining whether an individual has one of the NNMT haplotype pairs shown in Table 3. One way to accomplish this is to identify the phased sequence of nucleotides at PS1-PS3 for each copy of the individual's NNMT gene and assigning to the individual a NNMT haplotype pair that is consistent with each of the phased sequences, wherein the NNMT haplotype pair is selected from the group consisting of the NNMT haplotype pairs shown in Table 3. As described above, the identifying step does not necessarily require that each of PS1-PS3 be directly examined. As a result of linkage disequilibrium, typically only a subset of PS1-PS3 will need to be directly examined to assign to an individual a haplotype pair shown in Table 3.

When haplotyping both copies of the gene, the identifying step is preferably performed with each copy of the gene being placed in separate containers. However, it is also envisioned that if the two copies are labeled with different tags, or are otherwise separately distinguishable or identifiable, it could be possible in some cases to perform the method in the same container. For example, if first and second copies of the gene are labeled with different first and second fluorescent dyes, respectively,

and an allele-specific oligonucleotide labeled with yet a third different fluorescent dye is used to assay the polymorphic site(s), then detecting a combination of the first and third dyes would identify the polymorphism in the first gene copy while detecting a combination of the second and third dyes would identify the polymorphism in the second gene copy.

In both the genotyping and haplotyping methods, the identity of a nucleotide (or nucleotide pair) at a polymorphic site(s) may be determined by amplifying a target region(s) containing the polymorphic site(s) directly from one or both copies of the NNMT gene, or a fragment thereof, and the sequence of the amplified region(s) determined by conventional methods. It will be readily appreciated by the skilled artisan that only one nucleotide will be detected at a polymorphic site in individuals who are homozygous at that site, while two different nucleotides will be detected if the individual is heterozygous for that site. The polymorphism may be identified directly, known as positive-type identification, or by inference, referred to as negative-type identification. For example, where a SNP is known to be guanine and cytosine in a reference population, a site may be positively determined to be either guanine or cytosine for an individual homozygous at that site, or both guanine and cytosine, if the individual is heterozygous at that site. Alternatively, the site may be negatively determined to be not guanine (and thus cytosine/cytosine) or not cytosine (and thus guanine/guanine).

The target region(s) may be amplified using any oligonucleotide-directed amplification method, including but not limited to polymerase chain reaction (PCR) (U.S. Patent No. 4,965,188), ligase chain reaction (LCR) (Barany et al., *Proc. Natl. Acad. Sci. USA* 88:189-193, 1991; WO90/01069), and oligonucleotide ligation assay (OLA) (Landegren et al., *Science* 241:1077-1080, 1988). Other known nucleic acid amplification procedures may be used to amplify the target region including transcription-based amplification systems (U.S. Patent No. 5,130,238; EP 329,822; U.S. Patent No. 5,169,766, WO89/06700) and isothermal methods (Walker et al., *Proc. Natl. Acad. Sci. USA* 89:392-396, 1992).

A polymorphism in the target region may also be assayed before or after amplification using one of several hybridization-based methods known in the art. Typically, allele-specific oligonucleotides are utilized in performing such methods. The allele-specific oligonucleotides may be used as differently labeled probe pairs, with one member of the pair showing a perfect match to one variant of a target sequence and the other member showing a perfect match to a different variant. In some embodiments, more than one polymorphic site may be detected at once using a set of allele-specific oligonucleotides or oligonucleotide pairs. Preferably, the members of the set have melting temperatures within 5°C, and more preferably within 2°C, of each other when hybridizing to each of the polymorphic sites being detected.

Hybridization of an allele-specific oligonucleotide to a target polynucleotide may be performed with both entities in solution, or such hybridization may be performed when either the oligonucleotide or the target polynucleotide is covalently or noncovalently affixed to a solid support. Attachment may be mediated, for example, by antibody-antigen interactions, poly-L-Lys, streptavidin

or avidin-biotin, salt bridges, hydrophobic interactions, chemical linkages, UV cross-linking baking, etc. Allele-specific oligonucleotides may be synthesized directly on the solid support or attached to the solid support subsequent to synthesis. Solid-supports suitable for use in detection methods of the invention include substrates made of silicon, glass, plastic, paper and the like, which may be formed, for example, into wells (as in 96-well plates), slides, sheets, membranes, fibers, chips, dishes, and beads. The solid support may be treated, coated or derivatized to facilitate the immobilization of the allele-specific oligonucleotide or target nucleic acid.

The genotype or haplotype for the NNMT gene of an individual may also be determined by hybridization of a nucleic acid sample containing one or both copies of the gene, mRNA, cDNA or fragment(s) thereof, to nucleic acid arrays and subarrays such as described in WO 95/11995. The arrays would contain a battery of allele-specific oligonucleotides representing each of the polymorphic sites to be included in the genotype or haplotype.

The identity of polymorphisms may also be determined using a mismatch detection technique, including but not limited to the RNase protection method using riboprobes (Winter et al., *Proc. Natl. Acad. Sci. USA* 82:7575, 1985; Meyers et al., *Science* 230:1242, 1985) and proteins which recognize nucleotide mismatches, such as the *E. coli* mutS protein (Modrich, P. *Ann. Rev. Genet.* 25:229-253, 1991). Alternatively, variant alleles can be identified by single strand conformation polymorphism (SSCP) analysis (Orita et al., *Genomics* 5:874-879, 1989; Humphries et al., in Molecular Diagnosis of Genetic Diseases, R. Elles, ed., pp. 321-340, 1996) or denaturing gradient gel electrophoresis (DGGE) (Wartell et al., *Nucl. Acids Res.* 18:2699-2706, 1990; Sheffield et al., *Proc. Natl. Acad. Sci. USA* 86:232-236, 1989).

A polymerase-mediated primer extension method may also be used to identify the polymorphism(s). Several such methods have been described in the patent and scientific literature and include the "Genetic Bit Analysis" method (WO92/15712) and the ligase/polymerase mediated genetic bit analysis (U.S. Patent 5,679,524. Related methods are disclosed in WO91/02087, WO90/09455, WO95/17676, U.S. Patent Nos. 5,302,509, and 5,945,283. Extended primers containing a polymorphism may be detected by mass spectrometry as described in U.S. Patent No. 5,605,798. Another primer extension method is allele-specific PCR (Ruaño et al., *Nucl. Acids Res.* 17:8392, 1989; Ruaño et al., *Nucl. Acids Res.* 19, 6877-6882, 1991; WO 93/22456; Turki et al., *J. Clin. Invest.* 95:1635-1641, 1995). In addition, multiple polymorphic sites may be investigated by simultaneously amplifying multiple regions of the nucleic acid using sets of allele-specific primers as described in Wallace et al. (WO89/10414).

In addition, the identity of the allele(s) present at any of the novel polymorphic sites described herein may be indirectly determined by haplotyping or genotyping another polymorphic site that is in linkage disequilibrium with the polymorphic site that is of interest. Polymorphic sites in linkage disequilibrium with the presently disclosed polymorphic sites may be located in regions of the gene or in other genomic regions not examined herein. Detection of the allele(s) present at a polymorphic site

in linkage disequilibrium with the novel polymorphic sites described herein may be performed by, but is not limited to, any of the above-mentioned methods for detecting the identity of the allele at a polymorphic site.

In another aspect of the invention, an individual's NNMT haplotype pair is predicted from its NNMT genotype using information on haplotype pairs known to exist in a reference population. In its broadest embodiment, the haplotyping prediction method comprises identifying a NNMT genotype for the individual at two or more NNMT polymorphic sites described herein, accessing data containing NNMT haplotype pairs identified in a reference population, and assigning a haplotype pair to the individual that is consistent with the individual's NNMT genotype. In one embodiment, the reference haplotype pairs include the NNMT haplotype pairs shown in Table 3. The NNMT haplotype pair can be assigned by comparing the individual's genotype with the genotypes corresponding to the haplotype pairs known to exist in the general population or in a specific population group, and determining which haplotype pair is consistent with the genotype of the individual. In some embodiments, the comparing step may be performed by visual inspection (for example, by consulting Table 3). When the genotype of the individual is consistent with more than one haplotype pair, frequency data (such as that presented in Table 6) may be used to determine which of these haplotype pairs is most likely to be present in the individual. This determination may also be performed in some embodiments by visual inspection, for example by consulting Table 6. If a particular NNMT haplotype pair consistent with the genotype of the individual is more frequent in the reference population than others consistent with the genotype, then that haplotype pair with the highest frequency is the most likely to be present in the individual. In other embodiments, the comparison may be made by a computer-implemented algorithm with the genotype of the individual and the reference haplotype data stored in computer-readable formats. For example, as described in PCT/US01/12831, filed April 18, 2001, one computer-implemented algorithm to perform this comparison entails enumerating all possible haplotype pairs which are consistent with the genotype, accessing data containing NNMT haplotype pairs frequency data determined in a reference population to determine a probability that the individual has a possible haplotype pair, and analyzing the determined probabilities to assign a haplotype pair to the individual.

Generally, the reference population should be composed of randomly-selected individuals representing the major ethnogeographic groups of the world. A preferred reference population for use in the methods of the present invention comprises an approximately equal number of individuals from Caucasian, African-descent, Asian and Hispanic-Latino population groups with the minimum number of each group being chosen based on how rare a haplotype one wants to be guaranteed to see. For example, if one wants to have a q% chance of not missing a haplotype that exists in the population at a p% frequency of occurring in the reference population, the number of individuals (n) who must be sampled is given by $2n=\log(1-q)/\log(1-p)$ where p and q are expressed as fractions. A preferred reference population allows the detection of any haplotype whose frequency is at least 10% with about

99% certainty and comprises about 20 unrelated individuals from each of the four population groups named above. A particularly preferred reference population includes a 3-generation family representing one or more of the four population groups to serve as controls for checking quality of haplotyping procedures.

In a preferred embodiment, the haplotype frequency data for each ethnogeographic group is examined to determine whether it is consistent with Hardy-Weinberg equilibrium. Hardy-Weinberg equilibrium (D.L. Hartl et al., Principles of Population Genomics, Sinauer Associates (Sunderland, MA), 3rd Ed., 1997) postulates that the frequency of finding the haplotype pair H_1 / H_2 is equal to $p_{H-W}(H_1 / H_2) = 2p(H_1)p(H_2)$ if $H_1 \neq H_2$ and $p_{H-W}(H_1 / H_2) = p(H_1)p(H_2)$ if $H_1 = H_2$. A statistically significant difference between the observed and expected haplotype frequencies could be due to one or more factors including significant inbreeding in the population group, strong selective pressure on the gene, sampling bias, and/or errors in the genotyping process. If large deviations from Hardy-Weinberg equilibrium are observed in an ethnogeographic group, the number of individuals in that group can be increased to see if the deviation is due to a sampling bias. If a larger sample size does not reduce the difference between observed and expected haplotype pair frequencies, then one may wish to consider haplotyping the individual using a direct haplotyping method such as, for example, CLASPER System™ technology (U.S. Patent No. 5,866,404), single molecule dilution, or allele-specific long-range PCR (Michalotos-Beloin et al., *Nucleic Acids Res.* 24:4841-4843, 1996).

In one embodiment of this method for predicting a NNMT haplotype pair for an individual, the assigning step involves performing the following analysis. First, each of the possible haplotype pairs is compared to the haplotype pairs in the reference population. Generally, only one of the haplotype pairs in the reference population matches a possible haplotype pair and that pair is assigned to the individual. Occasionally, only one haplotype represented in the reference haplotype pairs is consistent with a possible haplotype pair for an individual, and in such cases the individual is assigned a haplotype pair containing this known haplotype and a new haplotype derived by subtracting the known haplotype from the possible haplotype pair. Alternatively, the haplotype pair in an individual may be predicted from the individual's genotype for that gene using reported methods (e.g., Clark et al. 1990 *Mol Bio Evol* 7:111-22 or WO 01/80156) or through a commercial haplotyping service such as offered by Genaissance Pharmaceuticals, Inc. (New Haven, CT). In rare cases, either no haplotypes in the reference population are consistent with the possible haplotype pairs, or alternatively, multiple reference haplotype pairs are consistent with the possible haplotype pairs. In such cases, the individual is preferably haplotyped using a direct molecular haplotyping method such as, for example, CLASPER System™ technology (U.S. Patent No. 5,866,404), SMD, or allele-specific long-range PCR (Michalotos-Beloin et al., *supra*).

The invention also provides a method for determining the frequency of a NNMT genotype, haplotype, or haplotype pair in a population. The method comprises, for each member of the

population, determining the genotype, haplotype or the haplotype pair for the novel NNMT polymorphic sites described herein, and calculating the frequency any particular genotype, haplotype, or haplotype pair is found in the population. The population may be e.g., a reference population, a family population, a same gender population, a population group, or a trait population (e.g., a group of individuals exhibiting a trait of interest such as a medical condition or response to a therapeutic treatment).

In one embodiment of the invention, NNMT haplotype frequencies in a trait population having a medical condition and a control population lacking the medical condition are used in a method of validating the NNMT protein as a candidate target for treating a medical condition predicted to be associated with NNMT activity. The method comprises comparing the frequency of each NNMT haplotype shown in Table 4 in the trait population and in a control population and making a decision whether to pursue NNMT as a target. It will be understood by the skilled artisan that the composition of the control population will be dependent upon the specific study and may be a reference population or it may be an appropriately matched population with regards to age, gender, and clinical symptoms for example. If at least one NNMT haplotype is present at a frequency in the trait population that is different from the frequency in the control population at a statistically significant level, a decision to pursue the NNMT protein as a target should be made. However, if the frequencies of each of the NNMT haplotypes are not statistically significantly different between the trait and control populations, a decision not to pursue the NNMT protein as a target is made. The statistically significant level of difference in the frequency may be defined by the skilled artisan practicing the method using any conventional or operationally convenient means known to one skilled in the art, taking into consideration that this level should help the artisan to make a rational decision about pursuing NNMT protein as a target. Any NNMT haplotype not present in a population is considered to have a frequency of zero. In some embodiments, each of the trait and controls populations may be comprised of different ethnogeographic origins, including but not limited to Caucasian, Hispanic Latino, African American, and Asian, while in other embodiments, the trait and reference population may be comprised of just one ethnogeographic origin.

In another embodiment of the invention, frequency data for NNMT haplotypes are determined in a population having a condition or disease predicted to be associated with NNMT activity and used in a method for screening for compounds targeting the NNMT protein to treat such condition or disease. In some embodiments, frequency data are determined in the population of interest for the NNMT haplotypes shown in Table 4. The frequency data for this population may be obtained by genotyping or haplotyping each individual in the population using one or more of the methods described above. The haplotypes for this population may be determined directly or, alternatively, by a predictive genotype to haplotype approach as described above. In another embodiment, the frequency data for this population are obtained by accessing previously determined frequency data, which may be in written or electronic form. For example, the frequency data may be present in a database that is

accessible by a computer. The NNMT isoforms corresponding to NNMT haplotypes occurring at a frequency greater than or equal to a desired frequency in this population are then used in screening for a compound, or compounds, that displays a desired agonist (enhancer) or antagonist (inhibitor) activity for each NNMT isoform. The desired frequency for the haplotypes might be chosen to be the frequency of the most frequent haplotype, greater than some cut-off value, such as 10% in the population, or the desired frequency might be determined by ranking the haplotypes by frequency and then choosing the frquency of the third most frequent haplotype as the cut-off value. Other methods for choosing a desired frequency are possible, such as choosing a frequency based on the desired market size for treatment with the compound. The desired level of agonist or antagonist level displayed in the screening process could be chosen to be greater than or equal to a cut-off value, such as activity levels in the top 10% of values determined. Embodiments may employ cell-free or cell-based screening assays known in the art. The compounds used in the screening assays may be from chemical compound libraries, peptide libraries and the like. The NNMT isoforms used in the screening assays may be free in solution, affixed to a solid support, or expressed in an appropriate cell line. In some embodiments, the condition or disease associated with NNMT activity is Parkinson's disease or cancer cachexia.

In another aspect of the invention, frequency data for NNMT genotypes, haplotypes, and/or haplotype pairs are determined in a reference population and used in a method for identifying an association between a trait and a NNMT genotype, haplotype, or haplotype pair. The trait may be any detectable phenotype, including but not limited to susceptibility to a disease or response to a treatment. In one embodiment, the method involves obtaining data on the frequency of the genotype(s), haplotype(s), or haplotype pair(s) of interest in a reference population as well as in a population exhibiting the trait. Frequency data for one or both of the reference and trait populations may be obtained by genotyping or haplotyping each individual in the populations using one or more of the methods described above. The haplotypes for the trait population may be determined directly or, alternatively, by a predictive genotype to haplotype approach as described above. In another embodiment, the frequency data for the reference and/or trait populations is obtained by accessing previously determined frequency data, which may be in written or electronic form. For example, the frequency data may be present in a database that is accessible by a computer. Once the frequency data is obtained, the frequencies of the genotype(s), haplotype(s), or haplotype pair(s) of interest in the reference and trait populations are compared. In a preferred embodiment, the frequencies of all genotypes, haplotypes, and/or haplotype pairs observed in the populations are compared. If a particular NNMT genotype, haplotype, or haplotype pair is different in the trait population than in the reference population to a statistically significant degree, then the trait is predicted to be associated with that NNMT genotype, haplotype or haplotype pair. Preferably, the NNMT genotype, haplotype, or haplotype pair being compared in the trait and reference populations is selected from the full-genotypes and full-haplotypes shown in Tables 3 and 4, or from sub-genotypes and sub-haplotypes

derived from these genotypes and haplotypes.

In a preferred embodiment of the method, the trait of interest is a clinical response exhibited by a patient to some therapeutic treatment, for example, response to a drug targeting or metabolized by NNMT or response to a therapeutic treatment for a medical condition. As used herein, "medical condition" includes but is not limited to any condition or disease manifested as one or more physical and/or psychological symptoms for which treatment is desirable, and includes previously and newly identified diseases and other disorders. As used herein the term "clinical response" means any or all of the following: a quantitative measure of the response, no response, and/or adverse response (i.e., side effects).

In order to deduce a correlation between clinical response to a treatment and a NNMT genotype, haplotype, or haplotype pair, it is necessary to obtain data on the clinical responses exhibited by a population of individuals who received the treatment, hereinafter the "clinical population". This clinical data may be obtained by analyzing the results of a clinical trial that has already been run and/or the clinical data may be obtained by designing and carrying out one or more new clinical trials. As used herein, the term "clinical trial" means any research study designed to collect clinical data on responses to a particular treatment, and includes but is not limited to phase I, phase II and phase III clinical trials. Standard methods are used to define the patient population and to enroll subjects.

It is preferred that the individuals included in the clinical population have been graded for the existence of the medical condition of interest. This is important in cases where the symptom(s) being presented by the patients can be caused by more than one underlying condition, and where treatment of the underlying conditions are not the same. An example of this would be where patients experience breathing difficulties that are due to either asthma or respiratory infections. If both sets were treated with an asthma medication, there would be a spurious group of apparent non-responders that did not actually have asthma. These people would affect the ability to detect any correlation between haplotype and treatment outcome. This grading of potential patients could employ a standard physical exam or one or more lab tests. Alternatively, grading of patients could use haplotyping for situations where there is a strong correlation between haplotype pair and disease susceptibility or severity.

The therapeutic treatment of interest is administered to each individual in the trial population and each individual's response to the treatment is measured using one or more predetermined criteria. It is contemplated that in many cases, the trial population will exhibit a range of responses and that the investigator will choose the number of responder groups (e.g., low, medium, high) made up by the various responses. In addition, the NNMT gene for each individual in the trial population is genotyped and/or haplotyped, which may be done before or after administering the treatment.

After both the clinical and polymorphism data have been obtained, correlations between individual response and NNMT genotype or haplotype content are created. Correlations may be produced in several ways. In one method, individuals are grouped by their NNMT genotype or

haplotype (or haplotype pair) (also referred to as a polymorphism group), and then the averages and standard deviations of clinical responses exhibited by the members of each polymorphism group are calculated.

These results are then analyzed to determine if any observed variation in clinical response between polymorphism groups is statistically significant. Statistical analysis methods which may be used are described in L.D. Fisher and G. vanBelle, "Biostatistics: A Methodology for the Health Sciences", Wiley-Interscience (New York) 1993. This analysis may also include a regression calculation of which polymorphic sites in the NNMT gene give the most significant contribution to the differences in phenotype. One regression model useful in the invention is described in WO 01/01218, entitled "Methods for Obtaining and Using Haplotype Data".

A second method for finding correlations between NNMT haplotype content and clinical responses uses predictive models based on error-minimizing optimization algorithms. One of many possible optimization algorithms is a genetic algorithm (R. Judson, "Genetic Algorithms and Their Uses in Chemistry" in Reviews in Computational Chemistry, Vol. 10, pp. 1-73, K. B. Lipkowitz and D. B. Boyd, eds. (VCH Publishers, New York, 1997). Simulated annealing (Press et al., "Numerical Recipes in C: The Art of Scientific Computing", Cambridge University Press (Cambridge) 1992, Ch. 10), neural networks (E. Rich and K. Knight, "Artificial Intelligence", 2nd Edition (McGraw-Hill, New York, 1991, Ch. 18), standard gradient descent methods (Press et al., *supra,* Ch. 10), or other global or local optimization approaches (see discussion in Judson, *supra*) could also be used. Preferably, the correlation is found using a genetic algorithm approach as described in WO 01/01218.

Correlations may also be analyzed using analysis of variation (ANOVA) techniques to determine how much of the variation in the clinical data is explained by different subsets of the polymorphic sites in the NNMT gene. As described in WO 01/01218, ANOVA is used to test hypotheses about whether a response variable is caused by or correlated with one or more traits or variables that can be measured (Fisher and vanBellé, *supra*, Ch. 10).

From the analyses described above, a mathematical model may be readily constructed by the skilled artisan that predicts clinical response as a function of NNMT genotype or haplotype content. Preferably, the model is validated in one or more follow-up clinical trials designed to test the model.

The identification of an association between a clinical response and a genotype or haplotype (or haplotype pair) for the NNMT gene may be the basis for designing a diagnostic method to determine those individuals who will or will not respond to the treatment, or alternatively, will respond at a lower level and thus may require more treatment, i.e., a greater dose of a drug. The diagnostic method will detect the presence in an individual of the genotype, haplotype or haplotype pair that is associated with the clinical response and may take one of several forms: for example, a direct DNA test (i.e., genotyping or haplotyping one or more of the polymorphic sites in the NNMT gene), a serological test, or a physical exam measurement. The only requirement is that there be a good correlation between the diagnostic test results and the underlying NNMT genotype or haplotype

that is in turn correlated with the clinical response. In a preferred embodiment, this diagnostic method uses the predictive haplotyping method described above.

Another embodiment of the invention comprises a method for reducing the potential for bias in a clinical trial of a candidate drug that targets or is metabolized by NNMT. Haplotyping one or both copies of the NNMT gene in those individuals participating in the trial will allow the pharmaceutical scientist conducting the clinical trial to assign each individual from the trial one of the haplotypes or haplotype pairs shown in Tables 4 and 3, respectively, in the NNMT gene. In one embodiment, the haplotypes may be determined directly, or alternatively, by a predictive genotype to haplotype approach as decribed above. In another embodiment, this can be accomplished by haplotyping individuals participating in a clinical trial by identifying, for example, in one or both copies of the individual's NNMT gene, the phased sequence of nucleotides present at each of PS1-PS3. Determining the NNMT haplotype or haplotype pair present in individuals participating in the clinical trial enables the pharmaceutical scientist to assign individuals possessing a specific haplotype or haplotype pair evenly to treatment and control groups. Typical clinical trials conducted may include, but are not limited to, Phase I, II, and III clinical trials. Each individual in the trial may produce a specific response to the candidate drug based upon the individual's haplotype or haplotype pair. To control for these differing drug responses in the trial and to reduce the potential for bias in the results that could be introduced by a larger frequency of an NNMT haplotype or haplotype pair in any particular treatment or control group due to random group assignment, each treatment and control group are assigned an even distribution (or equal numbers) of individuals having a particular NNMT haplotype or haplotype pair. To practice this method of the invention to reduce the potential for bias in a clinical trial, the pharmaceutical scientist requires no *a priori* knowledge of any effect a NNMT haplotype or haplotype pair may have on the results of the trial.

In another embodiment, the invention provides an isolated polynucleotide comprising a polymorphic variant of the NNMT gene or a fragment of the gene which contains at least one of the novel polymorphic sites described herein. The nucleotide sequence of a variant NNMT gene is identical to the reference genomic sequence for those portions of the gene examined, as described in the Examples below, except that it comprises a different nucleotide at one or more of the novel polymorphic sites PS1, PS2 and PS3. Similarly, the nucleotide sequence of a variant fragment of the NNMT gene is identical to the corresponding portion of the reference sequence except for having a different nucleotide at one or more of the novel polymorphic sites described herein. Thus, the invention specifically does not include polynucleotides comprising a nucleotide sequence identical to the reference sequence of the NNMT gene, which is defined by haplotype 3, (or other reported NNMT sequences) or to portions of the reference sequence (or other reported NNMT sequences), except for the haplotyping and genotyping oligonucleotides described above.

The location of a polymorphism in a variant NNMT gene or fragment is preferably identified by aligning its sequence against SEQ ID NO:1. The polymorphism is selected from the group

consisting of thymine at PS1, cytosine at PS2 and cytosine at PS3. In a preferred embodiment, the polymorphic variant comprises a naturally-occurring isogene of the NNMT gene which is defined by any one of haplotypes 1- 2 and 4 - 5 shown in Table 4 below.

Polymorphic variants of the invention may be prepared by isolating a clone containing the NNMT gene from a human genomic library. The clone may be sequenced to determine the identity of the nucleotides at the novel polymorphic sites described herein. Any particular variant or fragment thereof, that is claimed herein could be prepared from this clone by performing *in vitro* mutagenesis using procedures well-known in the art. Any particular NNMT variant or fragment thereof may also be prepared using synthetic or semi-synthetic methods known in the art.

NNMT isogenes, or fragments thereof, may be isolated using any method that allows separation of the two "copies" of the NNMT gene present in an individual, which, as readily understood by the skilled artisan, may be the same allele or different alleles. Separation methods include targeted *in vivo* cloning (TTVC) in yeast as described in WO 98/01573, U.S. Patent No. 5,866,404, and U.S. Patent No. 5,972,614. Another method, which is described in U.S. Patent No. 5,972,614, uses an allele specific oligonucleotide in combination with primer extension and exonuclease degradation to generate hemizygous DNA targets. Yet other methods are single molecule dilution (SMD) as described in Ruaño et al., *Proc. Natl. Acad. Sci.* 87:6296-6300, 1990; and allele specific PCR (Ruaño et al., 1989, *supra*; Ruaño et al., 1991, *supra*; Michalatos-Beloin et al., *supra*).

The invention also provides NNMT genome anthologies, which are collections of at least two NNMT isogenes found in a given population. The population may be any group of at least two individuals, including but not limited to a reference population, a population group, a family population, a clinical population, and a same gender population. A NNMT genome anthology may comprise individual NNMT isogenes stored in separate containers such as microtest tubes, separate wells of a microtitre plate and the like. Alternatively, two or more groups of the NNMT isogenes in the anthology may be stored in separate containers. Individual isogenes or groups of such isogenes in a genome anthology may be stored in any convenient and stable form, including but not limited to in buffered solutions, as DNA precipitates, freeze-dried preparations and the like. A preferred NNMT genome anthology of the invention comprises a set of isogenes defined by the haplotypes shown in Table 4 below.

An isolated polynucleotide containing a polymorphic variant nucleotide sequence of the invention may be operably linked to one or more expression regulatory elements in a recombinant expression vector capable of being propagated and expressing the encoded NNMT protein in a prokaryotic or a eukaryotic host cell. Examples of expression regulatory elements which may be used include, but are not limited to, the lac system, operator and promoter regions of phage lambda, yeast promoters, and promoters derived from vaccinia virus, adenovirus, retroviruses, or SV40. Other regulatory elements include, but are not limited to, appropriate leader sequences, termination codons, polyadenylation signals, and other sequences required for the appropriate transcription and subsequent

translation of the nucleic acid sequence in a given host cell. Of course, the correct combinations of expression regulatory elements will depend on the host system used. In addition, it is understood that the expression vector contains any additional elements necessary for its transfer to and subsequent replication in the host cell. Examples of such elements include, but are not limited to, origins of replication and selectable markers. Such expression vectors are commercially available or are readily constructed using methods known to those in the art (e.g., F. Ausubel et al., 1987, in "Current Protocols in Molecular Biology", John Wiley and Sons, New York, New York). Host cells which may be used to express the variant NNMT sequences of the invention include, but are not limited to, eukaryotic and mammalian cells, such as animal, plant, insect and yeast cells, and prokaryotic cells, such as *E. coli*, or algal cells as known in the art. The recombinant expression vector may be introduced into the host cell using any method known to those in the art including, but not limited to, microinjection, electroporation, particle bombardment, transduction, and transfection using DEAE-dextran, lipofection, or calcium phosphate (see e.g., Sambrook et al. (1989) in "Molecular Cloning. A Laboratory Manual", Cold Spring Harbor Press, Plainview, New York). In a preferred aspect, eukaryotic expression vectors that function in eukaryotic cells, and preferably mammalian cells, are used. Non-limiting examples of such vectors include vaccinia virus vectors, adenovirus vectors, herpes virus vectors, and baculovirus transfer vectors. Preferred eukaryotic cell lines include COS cells, CHO cells, HeLa cells, NIH/3T3 cells, and embryonic stem cells (Thomson, J. A. et al., 1998 *Science* 282:1145-1147). Particularly preferred host cells are mammalian cells.

As will be readily recognized by the skilled artisan, expression of polymorphic variants of the NNMT gene will produce NNMT mRNAs varying from each other at any polymorphic site retained in the spliced and processed mRNA molecules. These mRNAs can be used for the preparation of a NNMT cDNA comprising a nucleotide sequence which is a polymorphic variant of the NNMT reference coding sequence shown in Figure 2. Thus, the invention also provides NNMT mRNAs and corresponding cDNAs which comprise a nucleotide sequence that is identical to SEQ ID NO:2 (Fig. 2) (or its corresponding RNA sequence) for those regions of SEQ ID NO:2 that correspond to the examined portions of the NNMT gene (as described in the Examples below), except for having cytosine at a position corresponding to nucleotide 426. A particularly preferred polymorphic cDNA variant is A represented in Table 7. Fragments of these variant mRNAs and cDNAs are included in the scope of the invention, provided they contain the novel polymorphism described herein. The invention specifically excludes polynucleotides identical to previously identified NNMT mRNAs or cDNAs, and previously described fragments thereof. Polynucleotides comprising a variant NNMT RNA or DNA sequence may be isolated from a biological sample using well-known molecular biological procedures or may be chemically synthesized.

As used herein, a polymorphic variant of a NNMT gene, mRNA or cDNA fragment comprises at least one novel polymorphism identified herein and has a length of at least 10 nucleotides and may range up to the full length of the gene. Preferably, such fragments are between 100 and 3000

nucleotides in length, and more preferably between 200 and 2000 nucleotides in length, and most preferably between 200 and 750 nucleotides in length.

In describing the NNMT polymorphic sites identified herein, reference is made to the sense strand of the gene for convenience. However, as recognized by the skilled artisan, nucleic acid molecules containing the NNMT gene or cDNA may be complementary double stranded molecules and thus reference to a particular site on the sense strand refers as well to the corresponding site on the complementary antisense strand. Thus, reference may be made to the same polymorphic site on either strand and an oligonucleotide may be designed to hybridize specifically to either strand at a target region containing the polymorphic site. Thus, the invention also includes single-stranded polynucleotides which are complementary to the sense strand of the NNMT genomic, mRNA and cDNA variants described herein.

Polynucleotides comprising a polymorphic gene variant or fragment of the invention may be useful for therapeutic purposes. For example, where a patient could benefit from expression, or increased expression, of a particular NNMT protein isoform, an expression vector encoding the isoform may be administered to the patient. The patient may be one who lacks the NNMT isogene encoding that isoform or may already have at least one copy of that isogene.

In other situations, it may be desirable to decrease or block expression of a particular NNMT isogene. Expression of a NNMT isogene may be turned off by transforming a targeted organ, tissue or cell population with an expression vector that expresses high levels of untranslatable mRNA or antisense RNA for the isogene or fragment thereof. Alternatively, oligonucleotides directed against the regulatory regions (e.g., promoter, introns, enhancers, 3′ untranslated region) of the isogene may block transcription. Oligonucleotides targeting the transcription initiation site, e.g., between positions −10 and +10 from the start site are preferred. Similarly, inhibition of transcription can be achieved using oligonucleotides that base-pair with region(s) of the isogene DNA to form triplex DNA (see e.g., Gee et al. in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing Co., Mt. Kisco, N.Y., 1994). Antisense oligonucleotides may also be designed to block translation of NNMT mRNA transcribed from a particular isogene. It is also contemplated that ribozymes may be designed that can catalyze the specific cleavage of NNMT mRNA transcribed from a particular isogene.

The untranslated mRNA, antisense RNA or antisense oligonucleotides may be delivered to a target cell or tissue by expression from a vector introduced into the cell or tissue *in vivo* or *ex vivo*. Alternatively, such molecules may be formulated as a pharmaceutical composition for administration to the patient. Oligoribonucleotides and/or oligodeoxynucleotides intended for use as antisense oligonucleotides may be modified to increase stability and half-life. Possible modifications include, but are not limited to phosphorothioate or 2′ O-methyl linkages, and the inclusion of nontraditional bases such as inosine and queosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytosine, guanine, thymine, and uracil which are not as easily recognized by endogenous

nucleases.

Effect(s) of the polymorphisms identified herein on expression of NNMT may be investigated by various means known in the art, such as by *in vitro* translation of mRNA transcripts of the NNMT gene, cDNA or fragment thereof, or by preparing recombinant cells and/or nonhuman recombinant organisms, preferably recombinant animals, containing a polymorphic variant of the NNMT gene. As used herein, "expression" includes but is not limited to one or more of the following: transcription of the gene into precursor mRNA; splicing and other processing of the precursor mRNA to produce mature mRNA; mRNA stability; translation of the mature mRNA(s) into NNMT protein(s) (including effects of polymorphisms on codon usage and tRNA availability); and glycosylation and/or other modifications of the translation product, if required for proper expression and function.

To prepare a recombinant cell of the invention, the desired NNMT isogene, cDNA or coding sequence may be introduced into the cell in a vector such that the isogene, cDNA or coding sequence remains extrachromosomal. In such a situation, the gene will be expressed by the cell from the extrachromosomal location. In a preferred embodiment, the NNMT isogene, cDNA or coding sequence is introduced into a cell in such a way that it recombines with the endogenous NNMT gene present in the cell. Such recombination requires the occurrence of a double recombination event, thereby resulting in the desired NNMT gene polymorphism. Vectors for the introduction of genes both for recombination and for extrachromosomal maintenance are known in the art, and any suitable vector or vector construct may be used in the invention. Methods such as electroporation, particle bombardment, calcium phosphate co-precipitation and viral transduction for introducing DNA into cells are known in the art; therefore, the choice of method may lie with the competence and preference of the skilled practitioner. Examples of cells into which the NNMT isogene, cDNA or coding sequence may be introduced include, but are not limited to, continuous culture cells, such as COS, CHO, NIH/3T3, and primary or culture cells of the relevant tissue type, i.e., they express the NNMT isogene, cDNA or coding sequence. Such recombinant cells can be used to compare the biological activities of the different protein variants.

Recombinant nonhuman organisms, i.e., transgenic animals, expressing a variant NNMT gene, cDNA or coding sequence are prepared using standard procedures known in the art. Preferably, a construct comprising the variant gene, cDNA or coding sequence is introduced into a nonhuman animal or an ancestor of the animal at an embryonic stage, i.e., the one-cell stage, or generally not later than about the eight-cell stage. Transgenic animals carrying the constructs of the invention can be made by several methods known to those having skill in the art. One method involves transfecting into the embryo a retrovirus constructed to contain one or more insulator elements, a gene or genes (or cDNA or coding sequence) of interest, and other components known to those skilled in the art to provide a complete shuttle vector harboring the insulated gene(s) as a transgene, see e.g., U.S. Patent No. 5,610,053. Another method involves directly injecting a transgene into the embryo. A third method involves the use of embryonic stem cells. Examples of animals into which the NNMT

isogene, cDNA or coding sequences may be introduced include, but are not limited to, mice, rats, other rodents, and nonhuman primates (see "The Introduction of Foreign Genes into Mice" and the cited references therein, In: Recombinant DNA, Eds. J.D. Watson, M. Gilman, J. Witkowski, and M. Zoller; W.H. Freeman and Company, New York, pages 254-272). Transgenic animals stably expressing a human NNMT isogene, cDNA or coding sequence and producing the encoded human NNMT protein can be used as biological models for studying diseases related to abnormal NNMT expression and/or activity, and for screening and assaying various candidate drugs, compounds, and treatment regimens to reduce the symptoms or effects of these diseases.

An additional embodiment of the invention relates to pharmaceutical compositions affected by expression or function of a novel NNMT isogene described herein. The pharmaceutical composition may comprise any of the following active ingredients: a polynucleotide comprising one of these novel NNMT isogenes (or cDNAs or coding sequences); an antisense oligonucleotide directed against one of the novel NNMT isogenes, a polynucleotide encoding such an antisense oligonucleotide, or another compound which inhibits expression of a novel NNMT isogene described herein. Preferably, the composition contains the active ingredient in a therapeutically effective amount. By therapeutically effective amount is meant that one or more of the symptoms relating to disorders affected by expression or function of a novel NNMT isogene is reduced and/or eliminated. The composition also comprises a pharmaceutically acceptable carrier, examples of which include, but are not limited to, saline, buffered saline, dextrose, and water. Those skilled in the art may employ a formulation most suitable for the active ingredient, whether it is a polynucleotide, oligonucleotide, protein, peptide or small molecule antagonist. The pharmaceutical composition may be administered alone or in combination with at least one other agent, such as a stabilizing compound. Administration of the pharmaceutical composition may be by any number of routes including, but not limited to oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, intradermal, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal. Further details on techniques for formulation and administration may be found in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing Co., Easton, PA).

For any composition, determination of the therapeutically effective dose of active ingredient and/or the appropriate route of administration is well within the capability of those skilled in the art. For example, the dose can be estimated initially either in cell culture assays or in animal models. The animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans. The exact dosage will be determined by the practitioner, in light of factors relating to the patient requiring treatment, including but not limited to severity of the disease state, general health, age, weight and gender of the patient, diet, time and frequency of administration, other drugs being taken by the patient, and tolerance/response to the treatment.

Any or all analytical and mathematical operations involved in practicing the methods of the

present invention may be implemented by a computer. In addition, the computer may execute a program that generates views (or screens) displayed on a display device and with which the user can interact to view and analyze large amounts of information relating to the NNMT gene and its genomic variation, including chromosome location, gene structure, and gene family, gene expression data, polymorphism data, genetic sequence data, and clinical data population data (e.g., data on ethnogeographic origin, clinical responses, genotypes, and haplotypes for one or more populations). The NNMT polymorphism data described herein may be stored as part of a relational database (e.g., an instance of an Oracle database or a set of ASCII flat files). These polymorphism data may be stored on the computer's hard drive or may, for example, be stored on a CD-ROM or on one or more other storage devices accessible by the computer. For example, the data may be stored on one or more databases in communication with the computer via a network.

Preferred embodiments of the invention are described in the following examples. Other embodiments within the scope of the claims herein will be apparent to one skilled in the art from consideration of the specification or practice of the invention as disclosed herein. It is intended that the specification, together with the examples, be considered exemplary only, with the scope and spirit of the invention being indicated by the claims which follow the examples.

EXAMPLES

The Examples herein are meant to exemplify the various aspects of carrying out the invention and are not intended to limit the scope of the invention in any way. The Examples do not include detailed descriptions for conventional methods employed, such as in the performance of genomic DNA isolation, PCR and sequencing procedures. Such methods are well-known to those skilled in the art and are described in numerous publications, for example, Sambrook, Fritsch, and Maniatis, "Molecular Cloning: A Laboratory Manual", 2nd Edition, Cold Spring Harbor Laboratory Press, USA, (1989).

EXAMPLE 1

This example illustrates examination of various regions of the NNMT gene for polymorphic sites.

Amplification of Target Regions

The following target regions were amplified using either the PCR primers represented below or 'tailed' PCR primers, each of which includes a universal sequence forming a noncomplementary 'tail' attached to the 5′ end of each unique sequence in the PCR primer pairs. The universal 'tail' sequence for the forward PCR primers comprises the sequence 5′-TGTAAAACGACGGCCAGT-3′ (SEQ ID NO:19) and the universal 'tail' sequence for the reverse PCR primers comprises the sequence 5′-AGGAAACAGCTATGACCAT-3′ (SEQ ID NO:20). The nucleotide positions of the first and last

nucleotide of the forward and reverse primers for each region amplified are presented below and correspond to positions in SEQ ID NO:1 (Figure 1).

PCR Primer Pairs

Fragment No.	Forward Primer	Reverse Primer	PCR Product
Fragment 1	79-99	complement of 698-675	620 nt
Fragment 2	396-418	complement of 976-953	581 nt
Fragment 3	464-485	complement of 975-953	512 nt
Fragment 4	613-635	complement of 1160-1138	548 nt
Fragment 5	1924-1946	complement of 2415-2393	492 nt
Fragment 6	2540-2562	complement of 3063-3043	524 nt
Fragment 7	2659-2679	complement of 3313-3291	655 nt

These primer pairs were used in PCR reactions containing genomic DNA isolated from immortalized cell lines for each member of the Index Repository. The PCR reactions were carried out under the following conditions:

Reaction volume	= 10 µl
10 x Advantage 2 Polymerase reaction buffer (Clontech)	= 1 µl
100 ng of human genomic DNA	= 1 µl
10 mM dNTP	= 0.4 µl
Advantage 2 Polymerase enzyme mix (Clontech)	= 0.2 µl
Forward Primer (10 µM)	= 0.4 µl
Reverse Primer (10 µM)	= 0.4 µl
Water	= 6.6µl

Amplification profile:
97°C - 2 min. 1 cycle

97°C - 15 sec.
70°C - 45 sec.
72°C - 45 sec. } 10 cycles

97°C - 15 sec.
64°C - 45 sec.
72°C - 45 sec. } 35 cycles

Sequencing of PCR Products

The PCR products were purified using a Whatman/Polyfiltronics 100 µl 384 well unifilter plate essentially according to the manufacturers protocol. The purified DNA was eluted in 50 µl of distilled water. Sequencing reactions were set up using Applied Biosystems Big Dye Terminator chemistry essentially according to the manufacturers protocol. The purified PCR products were sequenced in both directions using either the primer sets represented below with the positions of their first and last nucleotide corresponding to positions in Figure 1, or the appropriate universal 'tail' sequence as a primer. Reaction products were purified by isopropanol precipitation, and run on an

Applied Biosystems 3700 DNA Analyzer.

Sequencing Primer Pairs

Fragment No.	Forward Primer	Reverse Primer
Fragment 1	150-171	complement of 671-652
Fragment 2	442-461	complement of 951-932
Fragment 3	495-514	complement of 912-893
Fragment 4	Tailed Seq.	
Fragment 5	Tailed Seq.	
Fragment 6	Tailed Seq.	
Fragment 7	2754-2773	complement of 3244-3224

Analysis of Sequences for Polymorphic Sites

Sequence information for a minimum of 80 humans was analyzed for the presence of polymorphisms using the Polyphred program (Nickerson et al., *Nucleic Acids Res.* 14:2745-2751, 1997). The presence of a polymorphism was confirmed on both strands. The polymorphisms and their locations in the NNMT reference genomic sequence (SEQ ID NO:1) are listed in Table 2 below.

Table 2. Polymorphic Sites Identified in the NNMT Gene

Polymorphic Site Number	Poly Id(a)	Nucleotide Position	Reference Allele	Variant Allele	CDS Variant Position	AA Variant
PS1	447356	394	A	T		
PS2	447360	928	T	C		
PS3	9736661	2696	T	C	426	D142D

(a) PolyId is a unique identifier assigned to each PS by Genaissance Pharmaceuticals, Inc.
(R) Reported previously.

EXAMPLE 2

This example illustrates analysis of the NNMT polymorphisms identified in the Index Repository for human genotypes and haplotypes.

The different genotypes containing these polymorphisms that were observed in unrelated members of the reference population are shown in Table 3 below, with the haplotype pair indicating the combination of haplotypes determined for the individual using the haplotype derivation protocol described below. In Table 3, homozygous positions are indicated by one nucleotide and heterozygous positions are indicated by two nucleotides. Missing nucleotides in any given genotype in Table 3 were inferred based on linkage disequilibrium and/or Mendelian inheritance.

Table 3 . Genotypes Observed for the NNMT Gene

Genotype Number	HAP Pair		Polymorphic Sites PS1	PS2	PS3
1	1	1	A	C	T
2	3	1	A	T/C	T
3	3	2	A	T	T/C
4	3	3	A	T	T
5	3	4	A/T	T/C	T
6	3	5	A/T	T	T
7	4	1	T/A	C	T
8	4	4	T	C	T
9	5	5	T	T	T

The haplotype pairs shown in Table 3 were estimated from the unphased genotypes using a computer-implemented algorithm for assigning haplotypes to unrelated individuals in a population sample, as described in WO 01/80156. In this method, haplotypes are assigned directly from individuals who are homozygous at all sites or heterozygous at no more than one of the variable sites. This list of haplotypes is then used to deconvolute the unphased genotypes in the remaining (multiply heterozygous) individuals. In the present analysis, the list of haplotypes was augmented with haplotypes obtained from two families (one three-generation Caucasian family and one two-generation African-American family).

By following this protocol, it was determined that the Index Repository examined herein and, by extension, the general population contains the 5 human NNMT haplotypes shown in Table 4 below, wherein each of the NNMT haplotypes comprises a 5′ – 3′ ordered sequence of 3 polymorphisms whose positions in SEQ ID NO:1 and identities are set forth in Table 4. In Table 4, the column labeled "Region Examined" provides the nucleotide positions in SEQ ID NO:1 corresponding to sequenced regions of the gene. The columns labeled "PS No." and "PS Position" provide the polymorphic site number designation (see Table 2) and the corresponding nucleotide position of this polymorphic site within SEQ ID NO:1 or SEQ ID NO:21. The columns beneath the "Haplotype Number" heading are labeled to provide a unique number designation for each NNMT haplotype.

Table 4. Haplotypes of the NNMT gene.

Region Examined(a)	PS No.(b)	PS Position(c)	Haplotype Number(d) 1	2	3	4	5
79-1160	1	394/30	A	A	A	T	T
79-1160	2	928/150	C	T	T	C	T
1924-2415	-	-	-	-	-	-	-
2540-3313	3	2696/270	T	C	T	T	T

(a) Region examined represents the nucleotide positions defining the start and stop positions within SEQ ID NO:1 of the regions sequenced;
(b) PS = polymorphic site;
(c) Position of PS within the indicated SEQ ID NO, with the 1st position number referring to SEQ ID NO:1 and the 2nd position number referring to SEQ ID NO:21, a modified version of SEQ ID NO:1 that comprises the context sequence of each polymorphic site, PS1-PS3, to facilitate electronic searching of the haplotypes;
(d) Alleles for NNMT haplotypes are presented 5′ to 3′ in each column.

SEQ ID NO:1 refers to Figure 1, with the two alternative allelic variants of each polymorphic site indicated by the appropriate nucleotide symbol. SEQ ID NO:21 is a modified version of SEQ ID NO:1 that shows the context sequence of each of PS1-PS3 in a uniform format to facilitate electronic searching of the NNMT haplotypes. For each polymorphic site, SEQ ID NO:21 contains a block of 60 bases of the nucleotide sequence encompassing the centrally-located polymorphic site at the 30th position, followed by 60 bases of unspecified sequence to represent that each polymorphic site is separated by genomic sequence whose composition is defined elsewhere herein.

Table 5 below shows the number of chromosomes characterized by a given NNMT haplotype for all unrelated individuals in the Index Repository for which haplotype data was obtained. The number of these unrelated individuals who have a given NNMT haplotype pair is shown in Table 6. In Tables 5 and 6, the "Total" column shows this frequency data for all of these unrelated individuals, while the other columns show the frequency data for these unrelated individuals categorized according to their self-identified ethnogeographic origin. Abbreviations used in Tables 5 and 6 are AF = African Descent, AS = Asian, CA = Caucasian, HL = Hispanic-Latino, and AM = Native American.

Table 5. Frequency of Observed NNMT Haplotypes In Unrelated Individuals

HAP No.	HAP ID	Total	CA	AF	AS	HL	AM
1	81445118	16	3	2	3	8	0
2	81445123	1	0	0	0	0	1
3	81445102	112	31	35	18	24	4
4	81445115	32	8	2	17	4	1
5	81445120	3	0	1	2	0	0

Table 6. Number of Observed NNMT Haplotype Pairs In Unrelated Individuals

HAP1	HAP2	Total	CA	AF	AS	HL	AM
1	1	1	0	0	0	1	0
3	1	11	2	2	2	5	0
3	2	1	0	0	0	0	1
3	3	42	11	15	6	9	1
3	4	15	7	2	4	1	1
3	5	1	0	1	0	0	0
4	1	3	1	0	1	1	0
4	4	7	0	0	6	1	0
5	5	1	0	0	1	0	0

The size and composition of the Index Repository were chosen to represent the genetic diversity across and within four major population groups comprising the general United States population. For example, as described in Table 1 above, this repository contains approximately equal sample sizes of African-descent, Asian-American, European-American, and Hispanic-Latino population groups. Almost all individuals representing each group had all four grandparents with the same ethnogeographic background. The number of unrelated individuals in the Index Repository provides a sample size that is sufficient to detect SNPs and haplotypes that occur in the general population with high statistical certainty. For instance, a haplotype that occurs with a frequency of 5% in the general population has a probability higher than 99.9% of being observed in a sample of 80 individuals from the general population. Similarly, a haplotype that occurs with a frequency of 10% in a specific population group has a 99% probability of being observed in a sample of 20 individuals from that population group. In addition, the size and composition of the Index Repository means that the relative frequencies determined therein for the haplotypes and haplotype pairs of the NNMT gene are likely to be similar to the relative frequencies of these NNMT haplotypes and haplotype pairs in the general U.S. population and in the four population groups represented in the Index Repository. The genetic diversity observed for the three Native Americans is presented because it is of scientific interest, but due to the small sample size it lacks statistical significance.

Each NNMT haplotype shown in Table 4 defines a NNMT isogene. The NNMT isogene defined by a given NNMT haplotype comprises the examined regions of SEQ ID NO:1 indicated in Table 4, with the corresponding ordered sequence of nucleotides occurring at each polymorphic site within the NNMT gene shown in Table 4 for that defining haplotype.

Each NNMT isogene defined by one of the haplotypes shown in Table 4 will further correspond to a particular NNMT coding sequence variant. Each of these NNMT coding sequence variants comprises the regions of SEQ ID NO:2 examined and is defined by the 5′ – 3′ ordered sequence of nucleotides occurring at each polymorphic site within the coding sequence of the NNMT gene, as shown in Table 7. In Table 7, the column labeled 'Region Examined' provides the nucleotide positions in SEQ ID NO:2 corresponding to sequenced regions of the gene; the columns labeled 'PS No.' and 'PS Position' provide the polymorphic site number designation (see Table 2) and the corresponding nucleotide position of this polymorphic site within SEQ ID NO:2. The columns beneath the 'Coding Sequence Number' heading are numbered to correspond to the haplotype number defining the NNMT isogene from which the coding sequence variant is derived. NNMT coding sequence variants that differ from the reference NNMT coding sequence are denoted in Table 7 by a letter (A, B, etc) identifying each unique novel coding sequence. The same letter at the top of more than one column denotes that a given novel coding sequence is present in multiple novel NNMT isogenes.

Table 7. Nucleotides Present at Polymorphic Sites Within the Observed NNMT Coding Sequences

Region Examined(a)	PS No.(b)	PS Position(c)	Coding Sequence Number(d) 1	2A	3	4	5
1-795	3	426	T	C	T	T	T

(a) Region examined represents the nucleotide positions in SEQ ID NO:2 defining the start and stop positions of the regions sequenced;
(b) PS = polymorphic site;
(c) Position of PS within SEQ ID NO:2;
(d) Alleles for NNMT coding sequences are presented 5′ to 3′ in each column. The number at the top of each column designates the haplotype number of the NNMT isogene from which the coding sequence is derived. NNMT coding sequences that differ from the reference are denoted in this table by a letter following the isogene number.

In view of the above, it will be seen that the several advantages of the invention are achieved and other advantageous results attained.

For any and all embodiments of the present invention discussed herein, in which a feature is described in terms of a Markush group or other grouping of alternatives, the inventors contemplate that such feature may also be described by, and that their invention specifically includes, any individual member or subgroup of members of such Markush group or other group.

As various changes could be made in the above methods and compositions without departing from the scope of the invention, it is intended that all matter contained in the above description and shown in the accompanying drawings shall be interpreted as illustrative and not in a limiting sense.

All references cited in this specification, including patents and patent applications, are hereby incorporated in their entirety by reference. The discussion of references herein is intended merely to summarize the assertions made by their authors and no admission is made that any reference

constitutes prior art. Applicants reserve the right to challenge the accuracy and pertinency of the cited references.

What is Claimed is:

1. A method for haplotyping the nicotinamide N-methyltransferase (NNMT) gene of an individual, which comprises identifying the phased sequence of nucleotides at PS1-PS3 for at least one copy of the individual's NNMT gene and assigning to the individual a NNMT haplotype that is consistent with the phased sequence, wherein the NNMT haplotype is selected from the group consisting of the NNMT haplotypes shown in the table immediately below:

PS No.(a)	PS Position(b)	Haplotype Number(c) 1	2	3	4	5
1	394	A	A	A	T	T
2	928	C	T	T	C	T
3	2696	T	C	T	T	T

(a) PS = polymorphic site; (b) Position of PS within SEQ ID NO:1;
(c) Alleles for haplotypes are presented 5′ to 3′ in each column.

2. A method for haplotyping the nicotinamide N-methyltransferase (NNMT) gene of an individual, which comprises identifying the phased sequence of nucleotides at PS1-PS3 for each copy of the individual's NNMT gene and assigning to the individual a NNMT haplotype pair that is consistent with each of the phased sequences, wherein the NNMT haplotype pair is selected from the group consisting of the NNMT haplotype pairs shown in the table immediately below:

PS No.(a)	PS Position(b)	Haplotype Pair(c)(Part 1) 1/1	3/1	3/2	3/3	3/4	3/5	4/1	4/4
1	394	A/A	A/A	A/A	A/A	A/T	A/T	T/A	T/T
2	928	C/C	T/C	T/T	T/T	T/C	T/T	C/C	C/C
3	2696	T/T	T/T	T/C	T/T	T/T	T/T	T/T	T/T

PS No.(a)	PS Position(b)	Haplotype Pair(c)(Part 2) 5/5
1	394	T/T
2	928	T/T
3	2696	T/T

(a) PS = polymorphic site; (b) Position of PS in SEQ ID NO:1;
(c) Haplotype pairs are represented as 1st haplotype/2nd haplotype; with alleles of each haplotype shown 5′ to 3′ as 1st polymorphism/2nd polymorphism in each column.

3. A method for genotyping the nicotinamide N-methyltransferase (NNMT) gene of an individual, comprising determining for the two copies of the NNMT gene present in the individual the identity of the nucleotide pair at one or more polymorphic sites (PS) selected from the group consisting of PS1, PS2 and PS3, wherein the one or more polymorphic sites (PS) have the position and alternative alleles shown in SEQ ID NO:1.

4. The method of claim 3, wherein the determining step comprises:

(a) isolating from the individual a nucleic acid mixture comprising both copies of the NNMT gene, or a fragment thereof, that are present in the individual;

(b) amplifying from the nucleic acid mixture a target region containing one of the selected polymorphic sites;

(c) hybridizing a primer extension oligonucleotide to one allele of the amplified target region, wherein the oligonucleotide is designed for genotyping the selected polymorphic site in the target region;

(d) performing a nucleic acid template-dependent, primer extension reaction on the hybridized oligonucleotide in the presence of at least one terminator of the reaction, wherein the terminator is complementary to one of the alternative nucleotides present at the selected polymorphic site; and

(e) detecting the presence and identity of the terminator in the extended oligonucleotide.

5. The method of claim 3, which comprises determining for the two copies of the NNMT gene present in the individual the identity of the nucleotide pair at each of PS1-PS3.

6. A method for haplotyping the nicotinamide N-methyltransferase (NNMT) gene of an individual which comprises determining, for one copy of the NNMT gene present in the individual, the identity of the nucleotide at two or more polymorphic sites (PS) selected from the group consisting of PS1, PS2 and PS3, wherein the selected PS have the position and alternative alleles shown in SEQ ID NO:1.

7. The method of claim 6, wherein the determining step comprises:

(a) isolating from the individual a nucleic acid sample containing only one of the two copies of the NNMT gene, or a fragment thereof, that is present in the individual;

(b) amplifying from the nucleic acid sample a target region containing one of the selected polymorphic sites;

(c) hybridizing a primer extension oligonucleotide to one allele of the amplified target region, wherein the oligonucleotide is designed for haplotyping the selected polymorphic site in the target region;

(d) performing a nucleic acid template-dependent, primer extension reaction on the hybridized oligonucleotide in the presence of at least one terminator of the reaction, wherein the terminator is complementary to one of the alternative nucleotides present at the selected polymorphic site; and

(e) detecting the presence and identity of the terminator in the extended oligonucleotide.

8. A method for predicting a haplotype pair for the nicotinamide N-methyltransferase (NNMT) gene of an individual comprising:

(a) identifying a NNMT genotype for the individual, wherein the genotype comprises the nucleotide pair at two or more polymorphic sites (PS) selected from the group consisting of PS1, PS2 and PS3, wherein the selected PS have the position and alternative alleles shown in SEQ ID NO:1;

(b) comparing the genotype to the haplotype pair data set forth in the table immediately below; and

(c) determining which haplotype pair is consistent with the genotype of the individual and

with the haplotype pair data

PS No.(a)	PS Position(b)	Haplotype Pair(c)(Part 1) 1/1	3/1	3/2	3/3	3/4	3/5	4/1	4/4
1	394	A/A	A/A	A/A	A/A	A/T	A/T	T/A	T/T
2	928	C/C	T/C	T/T	T/T	T/C	T/T	C/C	C/C
3	2696	T/T	T/T	T/C	T/T	T/T	T/T	T/T	T/T

PS No.(a)	PS Position(b)	Haplotype Pair(c)(Part 2) 5/5
1	394	T/T
2	928	T/T
3	2696	T/T

(a) PS = polymorphic site; (b) Position of PS in SEQ ID NO:1;
(c) Haplotype pairs are represented as 1st haplotype/2nd haplotype; with alleles of each haplotype shown 5′ to 3′ as 1st polymorphism/2nd polymorphism in each column.

9. The method of claim 8, wherein the identified genotype of the individual comprises the nucleotide pair at each of PS1-PS3, which have the position and alternative alleles shown in SEQ ID NO:1.

10. A method for identifying an association between a trait and at least one haplotype or haplotype pair of the nicotinamide N-methyltransferase (NNMT) gene which comprises comparing the frequency of the haplotype or haplotype pair in a population exhibiting the trait with the frequency of the haplotype or haplotype pair in a reference population, wherein the haplotype is selected from haplotypes 1-5 shown in the table presented immediately below:

PS No.(a)	PS Position(b)	Haplotype Number(c) 1	2	3	4	5
1	394	A	A	A	T	T
2	928	C	T	T	C	T
3	2696	T	C	T	T	T

(a) PS = polymorphic site; (b) Position of PS within SEQ ID NO:1;
(c) Alleles for haplotypes are presented 5′ to 3′ in each column;

and wherein the haplotype pair is selected from the haplotype pairs shown in the table immediately below:

PS No.(a)	PS Position(b)	Haplotype Pair(c)(Part 1) 1/1	3/1	3/2	3/3	3/4	3/5	4/1	4/4
1	394	A/A	A/A	A/A	A/A	A/T	A/T	T/A	T/T
2	928	C/C	T/C	T/T	T/T	T/C	T/T	C/C	C/C
3	2696	T/T	T/T	T/C	T/T	T/T	T/T	T/T	T/T

PS No.(a)	PS Position(b)	Haplotype Pair(c)(Part 2) 5/5
1	394	T/T
2	928	T/T
3	2696	T/T

(a) PS = polymorphic site; (b) Position of PS in SEQ ID NO:1;
(c) Haplotype pairs are represented as 1st haplotype/2nd haplotype; with alleles of each haplotype shown 5′ to 3′ as 1st polymorphism/2nd polymorphism in each column;

wherein a statistically significant different frequency of the haplotype or haplotype pair in the trait population than in the reference population indicates the trait is associated with the haplotype or haplotype pair.

11. The method of claim 10, wherein the trait is a clinical response to a drug that binds to or is metabolized by NNMT.

12. The method of claim 11, which further comprises designing a diagnostic method for determining those individuals who will exhibit the clinical response, wherein the method detects the presence in an individual of the haplotype or haplotype pair associated with the clinical response.

13. The method of claim 10, wherein the trait is a clinical response to a drug for treating a condition or disease predicted to be associated with NNMT activity.

14. The method of claim 13, which further comprises designing a diagnostic method for determining those individuals who will exhibit the clinical response, wherein the method detects the presence in an individual of the haplotype or haplotype pair associated with the clinical response.

15. A method for reducing the potential for bias in a clinical trial of a candidate drug that binds to or is metabolized by NNMT, the method comprising determining which of the NNMT haplotypes or NNMT haplotype pairs shown in the tables immediately below are present in each individual that is participating in the trial; and assigning each individual to a treatment group or a control group to produce an even distribution of each of the determined NNMT haplotypes or NNMT haplotype pairs in the treatment group and the control group,

PS	PS	Haplotype Number(c)				
No.(a)	Position(b)	1	2	3	4	5
1	394	A	A	A	T	T
2	928	C	T	T	C	T
3	2696	T	C	T	T	T

(a) PS = polymorphic site; (b) Position of PS within SEQ ID NO:1;
(c) Alleles for haplotypes are presented 5′ to 3′ in each column;

PS	PS	Haplotype Pair(c)(Part 1)							
No.(a)	Position(b)	1/1	3/1	3/2	3/3	3/4	3/5	4/1	4/4
1	394	A/A	A/A	A/A	A/A	A/T	A/T	T/A	T/T
2	928	C/C	T/C	T/T	T/T	T/C	T/T	C/C	C/C
3	2696	T/T	T/T	T/C	T/T	T/T	T/T	T/T	T/T

PS	PS	Haplotype Pair(c)(Part 2)
No.(a)	Position(b)	5/5
1	394	T/T
2	928	T/T
3	2696	T/T

(a) PS = polymorphic site; (b) Position of PS in SEQ ID NO:1;
(c) Haplotype pairs are represented as 1st haplotype/2nd haplotype; with alleles of each haplotype shown 5′ to 3′ as 1st polymorphism/2nd polymorphism in each column.

16. An isolated polynucleotide comprising a nucleotide sequence selected from the group consisting of:

(a) a first nucleotide sequence which comprises a nicotinamide N-methyltransferase (NNMT) isogene, wherein the NNMT isogene is selected from the group consisting of isogenes 1-2 and 4 - 5 shown in the table immediately below and wherein each of the isogenes comprises the regions of SEQ ID NO:1 shown in the table immediately below, except where substituted by the corresponding sequence of polymorphisms whose positions and identities are set forth in the table immediately below; and

Region	PS	PS	Isogene Number(d)			
Examined(a)	No.(b)	Position(c)	1	2	4	5
79-1160	1	394	A	A	T	T
79-1160	2	928	C	T	C	T
1924-2415	-	-	-	-	-	-
2540-3313	3	2696	T	C	T	T

(a) Region examined represents the nucleotide positions defining the start and stop positions within the 1st SEQ ID NO of the sequenced region;
(b) PS = polymorphic site;
(c) Position of PS in SEQ ID NO:1;
(d) Alleles for isogenes are presented 5′ to 3′ in each column;

(b) a second nucleotide sequence which is complementary to the first nucleotide sequence.

17. The isolated polynucleotide of claim 16, which is a DNA molecule and comprises both the first and second nucleotide sequences and further comprises expression regulatory elements operably linked to the first nucleotide sequence.

18. A recombinant nonhuman organism transformed or transfected with the isolated polynucleotide of claim 17, wherein the organism expresses a NNMT protein that is encoded by the first nucleotide sequence.

19. The recombinant nonhuman organism of claim 18, which is a transgenic animal.

20. An isolated fragment of a nicotinamide N-methyltransferase (NNMT) isogene, wherein the fragment comprises at least 10 nucleotides in one of the regions of SEQ ID NO:1 shown in the table immediately below and wherein the fragment comprises one or more polymorphisms selected from the group consisting of thymine at PS1, cytosine at PS2 and cytosine at PS3, wherein the selected polymorphism has the position set forth in the table immediately below:

Region	PS	PS	Isogene Number(d)			
Examined(a)	No.(b)	Position(c)	1	2	4	5
79-1160	1	394	A	A	T	T
79-1160	2	928	C	T	C	T
1924-2415	-	-	-	-	-	-
2540-3313	3	2696	T	C	T	T

(a) Region examined represents the nucleotide positions defining the start and stop positions within SEQ ID NO:1 of the regions sequenced; (b) PS = polymorphic site; (c) Position of PS within SEQ ID NO:1; (d) Alleles for NNMT isogenes are presented 5′ to 3′ in each column.

21. The isolated fragment of claim 20, wherein the fragment has a length between 200 and 750

nucleotides.

22. An isolated polynucleotide comprising a coding sequence variant for a NNMT isogene, wherein the coding sequence variant comprises nucleotides 1-795 in SEQ ID NO:2, and wherein the selected coding sequence variant further comprises cytosine at a position corresponding to nucleotide 426 in SEQ ID NO:2.

23. A recombinant nonhuman organism transformed or transfected with the isolated polynucleotide of claim 22, wherein the organism expresses a nicotinamide N-methyltransferase (NNMT) protein that is encoded by the polymorphic variant sequence.

24. The recombinant nonhuman organism of claim 23, which is a transgenic animal.

25. An isolated fragment of a NNMT coding sequence, wherein the fragment comprises cytosine at a position corresponding to nucleotide 426 in SEQ ID NO:2.

26. The isolated fragment of claim 25, wherein the fragment has a length between 200 and 750 nucleotides.

27. A method for validating the NNMT protein as a candidate target for treating a medical condition predicted to be associated with NNMT activity, the method comprising:

(a) comparing the frequency of each of the NNMT haplotypes in the table shown immediately below between first and second populations, wherein the first population is a group of individuals having the medical condition and the second population is a group of individuals lacking the medical condition; and

(b) making a decision whether to pursue NNMT as a target for treating the medical condition; wherein if at least one of the NNMT haplotypes is present in a frequency in the first population that is different from the frequency in the second population at a statistically significant level, then the decision is to pursue the NNMT protein as a target and if none of the NNMT haplotypes are seen in a different frequency, at a statistically significant level, between the first and second populations, then the decision is to not pursue the NNMT protein as a target.

PS	PS	Haplotype Number(c)				
No.(a)	Position(b)	1	2	3	4	5
1	394	A	A	A	T	T
2	928	C	T	T	C	T
3	2696	T	C	T	T	T

(a) PS = polymorphic site; (b) Position of PS within SEQ ID NO:1;
(c) Alleles for haplotypes are presented 5′ to 3′ in each column.

28. The method of claim 27, wherein the condition or disease is Parkinson's disease or cancer cachexia.

29. An isolated oligonucleotide designed for detecting a polymorphism in the nicotinamide N-methyltransferase (NNMT) gene at a polymorphic site (PS) selected from the group consisting of PS1, PS2 and PS3, wherein the selected PS have the position and alternative alleles shown in SEQ ID NO:1.

30. The isolated oligonucleotide of claim 29, which is an allele-specific oligonucleotide that specifically hybridizes to an allele of the NNMT gene at a region containing the polymorphic site.

31. The allele-specific oligonucleotide of claim 30, which comprises a nucleotide sequence selected from the group consisting of SEQ ID NOS:4-6, the complements of SEQ ID NOS:4-6, and SEQ ID NOS:7-12.

32. The isolated oligonucleotide of claim 29, which is a primer-extension oligonucleotide.

33. The primer-extension oligonucleotide of claim 32, which comprises a nucleotide sequence selected from the group consisting of SEQ ID NOS:13-18.

34. A kit for haplotyping or genotyping the nicotinamide N-methyltransferase (NNMT) gene of an individual, which comprises a set of oligonucleotides designed to haplotype or genotype each of polymorphic sites (PS) PS1, PS2 and PS3, wherein the selected PS have the position and alternative alleles shown in SEQ ID NO:1.

35. A computer system for storing and analyzing polymorphism data for the nicotinamide N-methyltransferase gene, comprising:

(a) a central processing unit (CPU);

(b) a communication interface;

(c) a display device;

(d) an input device; and

(e) a database containing the polymorphism data;

wherein the polymorphism data comprises the haplotypes set forth in the table immediately below:

PS	PS	Haplotype Number(c)				
No.(a)	Position(b)	1	2	3	4	5
1	394	A	A	A	T	T
2	928	C	T	T	C	T
3	2696	T	C	T	T	T

(a) PS = polymorphic site; (b) Position of PS within SEQ ID NO:1;
(c) Alleles for haplotypes are presented 5′ to 3′ in each column;

the haplotype pairs set forth in the table immediately below:

PS No.(a)	PS Position(b)	Haplotype Pair(c)(Part 1) 1/1	3/1	3/2	3/3	3/4	3/5	4/1	4/4
1	394	A/A	A/A	A/A	A/A	A/T	A/T	T/A	T/T
2	928	C/C	T/C	T/T	T/T	T/C	T/T	C/C	C/C
3	2696	T/T	T/T	T/C	T/T	T/T	T/T	T/T	T/T

PS No.(a)	PS Position(b)	Haplotype Pair(c)(Part 2) 5/5
1	394	T/T
2	928	T/T
3	2696	T/T

(a) PS = polymorphic site; (b) Position of PS in SEQ ID NO:1;
(c) Haplotype pairs are represented as 1st haplotype/2nd haplotype; with alleles of each haplotype shown 5′ to 3′ as 1st polymorphism/2nd polymorphism in each column;

or the frequency data in Tables 5 and 6.

36. A genome anthology for the nicotinamide N-methyltransferase (NNMT) gene which comprises two or more NNMT isogenes selected from the group consisting of isogenes 1-5 shown in the table immediately below, and wherein each of the isogenes comprises the regions of SEQ ID NO:1 shown in the table immediately below and wherein each of the isogenes 1-5 is further defined by the corresponding sequence of polymorphisms whose positions and identities are set forth in the table immediately below:

Region Examined(a)	PS No.(b)	PS Position(c)	Isogene Number(d) 1	2	3	4	5
79-1160	1	394	A	A	A	T	T
79-1160	2	928	C	T	T	C	T
1924-2415	-	-	-	-	-	-	-
2540-3313	3	2696	T	C	T	T	T

(a) Region examined represents the nucleotide positions defining the start and stop positions within SEQ ID NO:1 of the regions sequenced; (b) PS = polymorphic site; (c) Position of PS within SEQ ID NO:1; (d) Alleles for NNMT isogenes are presented 5′ to 3′ in each column.

POLYMORPHISMS IN THE NNMT GENE

```
CAGACACTGG GTCATGGCAG TGGTCGGTGA AGCTGCAGTT GCCTAGGGCA
GGGATGGAGA GAGAGTCTGG GCATGAGGAG AGGGTCTCGG GATGTTTGGC 100
TGGACTAGAT TTTACAGAAA GCCTTATCCA GGCTTTTAAA ATTACTCTTT
CCAGACTTCA TCTGAGACTC CTTCTTCAGC CAACATTCCT TAGCCCTGAA 200
TACATTTCCT ATCCTCATCT TTCCCTTCTT TTTTTTCCTT TCTTTTACAT
GTTTAAATTT AAACCATTCT TCGTGACCCC TTTTCTTGGG AGATTCATGG 300
CAAGAACGAG AAGAATGATG GTGCTTGTTA GGGGATGTCC TGTCTCTCTG
AACTTTGGGG TCCTATGCAA TAAATAATTT TCCTGACGAG CTCAAGTGCT 400
                                                    T
CCCTCTGGTC TACAATCCCT GGCGGCTGGC CTTCATCCCT TGGGCAAGCA
TTGCATACAG CTCATGGCCC TCCCTCTACC ATACCCTCCA CCCCCGTTCG 500
CCTAAGCTCC CTTCTCCGGG AATTTCATCA TTTCCTAGAA CAGCCAGAAC
ATTTGTGGTC TATTTCTCTG TTAGTGTTTA ACCAACCATC TGTTCTAAAA 600
GAAGGGCTGA ACTGATGGAA GGAATGCTGT TAGCCTGAGA CTCAGGAAGA
CAACTTCTGC AGGGTCACTC CCTGGCTTCT GGAGGAAAGA GAAGGAGGGC 700
AGTGCTCCAG TGGTACAGAA GTGAGACATA ATGGAATCAG GCTTCACCTC
[EXON 1: 731..
CAAGGACACC TATCTAAGCC ATTTTAACCC TCGGGATTAC CTAGAAAAAT 800
ATTACAAGTT TGGTTCTAGG CACTCTGCAG AAAGCCAGAT TCTTAAGCAC
CTTCTGAAAA ATCTTTTCAA GATATTCTGC CTAGGTAAGT CTGTTGTCTG 900
.. 884]
CATGTCTCCC CACTAATGTG AGTCATATAG ATGGAGTCTC AGGGCACGAC
                              C
TGGGTTTTGT GTCTCTCGTT GTTGCTTCAC AGCCCTTTTG GCATCACCCA 1000
TTTATTTAAC TAGGATAAAA ACGAATATTG GTATAGCGAT TCCACAGTTT
ACAAAGTGCT TTTGTATCCA CTGTCTCACT TGATCAAGCA AAAGGAAACC 1100
AGAGGACCGG AGTGCTGTCC TGAGTCTACC TTGATTTGCT AGGCGACTTG
AGGGAGACTT TTAGCCTCAA AGGGCCACTT AAGTGGAAAT TCTAAAACAG 1200
TACCTATTCT GATCCTAACT CAAGGGAATG CTGTGAATAT GCATGAGATA
AAGACCTCCC AATATATGAA GAACTGGGTG ATTTTGGGAG AAAGACATTA 1300
TATACTCAAT TTCTTTTTTA ATTAACTTTC CTTGAAAGTA TTGCTTAATA
GTTTTTACAT TCTCCATGTA ACAGACTTTC TGGATCTGGT CTTCAGTCTG 1400
TACACCAGAT GTAGATCTTT TTTACCTTCT CCTAGACCTT AAAATTCCTG
GCAACATGCC TCCACCCTGG ATTGGGGAAT AAAAAATGAA AAGTTTTTTT 1500
TTTTTCTTTT TGACTTTAAA TTTTATTAAA GTTTGAGGTT TTTCAAACTG
ATGTGCTTTA TTTAAAATCC AAGTGAGACA TTTTTAGTCT TTTTGATATT 1600
TATATTTTCT TTGTCACTAT GATGTAAATT ACAGGGATTT GGGGGAAAAA
TGGGATTTTT TTTTTTTTTT TGGAGATATA GATCTCACTC TGTTTCCTAG 1700
GCTGGATGGA GTGCAGGGAT GTGATCACAG CTCATCATAG TCTCGAACTC
CTGGACTCAA GGGATCCTTC TGCCTCAGCC TCTCCAATAA CTAGGTCTTC 1800
AGGCACACGC CACCATGCCT AGCTAATTTT AAAATTTTTT TGTAGAGATG
AGGTCTCACT ATGTTGTCCA GGCTGGTCTC ATCCTCCAGG CCTCAAGTGA 1900
TCCTCCTGCC TTGGCCTTCC AGAGTATTGG GACTGTAGGC ATGAGCCACT
GTGCCTGGCC CAGAAAAGAT GTTTTAAAAA AACATTTTGA GGGAAAAGTT 2000
GTGAACAGTA GTGGTCTGTC TTTGAGGATC GCCAGCACAG TCCCAGGGAA
GACAATGTAA ATTTGACTCT GCCCACTGCC ATGAGATGCC TGATCTCTCC 2100
TCTTTGTTCC TCCCACTAAT CCAGACGGTG TGAAGGGAGA CCTGCTGATT
[EXON 2: 2125..
GACATCGGCT CTGGCCCCAC TATCTATCAG CTCCTCTCTG CTTGTGAATC 2200
CTTTAAGGAG ATCGTCGTCA CTGACTACTC AGACCAGAAC CTGCAGGAGC
TGGAGAAGTG GCTGAAGAAA GAGCCAGAGG CCTTTGACTG GTCCCCAGTG 2300
```

FIGURE 1A

```
GTGACCTATG TGTGTGATCT TGAAGGGAAC AGGTAGAGAA ACTGGTGTCT
.. 2332]
ACTTCTTGGC TTTTGAAGGT ACCTGAGTGA TGGTTGGCAA AAGCAACAGA 2400
CAGATAGGGA CCAAAGAGAA ATCCAAATGG AGNNNNNNNN NNNNNNNNNN
NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN 2500
NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNCACATCTG GGACAATACG
[EXON 3: 2533..
GCCATTTTTA GCCTTGACCC AAGAGATCTG GGTTCCCCAT GACTGGAGTG 2600
GAAAACAATG TCTGTGGGTT TGTGTTTTTC AGAGTCAAGG GTCCAGAGAA
GGAGGAGAAG TTGAGACAGG CGGTCAAGCA GGTGCTGAAG TGTGATGTGA 2700
                                                    C
CTCAGAGCCA GCCACTGGGG GCCGTCCCCT TACCCCCGGC TGACTGCGTG
CTCAGCACAC TGTGTCTGGA TGCCGCCTGC CCAGACCTCC CCACCTACTG 2800
CAGGGCGCTC AGGAACCTCG GCAGCCTACT GAAGCCAGGG GGCTTCCTGG
TGATCATGGA TGCGCTCAAG AGCAGCTACT ACATGATTGG TGAGCAGAAG 2900
TTCTCCAGCC TCCCCCTGGG CCGGGAGGCA GTAGAGGCTG CTGTGAAAGA
GGCTGGCTAC ACAATCGAAT GGTTTGAGGT GATCTCGCAA AGTTATTCTT 3000
.. 2965]
CCACCATGGC CAACAACGAA GGACTTTTCT CCCTGGTGGC GAGGAAGCTG
AGCAGACCCC TGTGATGCCT GTGACCTCAA TTAAAGCAAT TCCTTTGACC 3100
TGTCCAGTTG ACTTTAGTCC TTGTTTCTAA CTGCCAAGTC ATGTGCTGAG
TAGAGGCTCA GTGGTTGGGG CCCAATGGTT CATCTAGGAC GGGACTAGAG 3200
AGGTCAGTCT ACAAGCAATC CATTGACCAC TTACTTGGTG CTGCACACAA
ATGTTGGTGC TATGGGACCC AAAGATGAGC AATTAGTATT CCAGTCTTCA 3300
TTGCCTGTGC TTACAAAAGA AGACCTCACT TCCCTAAACA TCTAGTTATG
GAGGTTCAAG CCCGTACCTG CCTACAGAGA AGTGT                 3385
```

FIGURE 1B

POLYMORPHISMS IN THE CODING SEQUENCE OF NNMT

```
ATGGAATCAG GCTTCACCTC CAAGGACACC TATCTAAGCC ATTTTAACCC
TCGGGATTAC CTAGAAAAAT ATTACAAGTT TGGTTCTAGG CACTCTGCAG 100
AAAGCCAGAT TCTTAAGCAC CTTCTGAAAA ATCTTTTCAA GATATTCTGC
CTAGACGGTG TGAAGGGAGA CCTGCTGATT GACATCGGCT CTGGCCCCAC 200
TATCTATCAG CTCCTCTCTG CTTGTGAATC CTTTAAGGAG ATCGTCGTCA
CTGACTACTC AGACCAGAAC CTGCAGGAGC TGGAGAAGTG GCTGAAGAAA 300
GAGCCAGAGG CCTTTGACTG GTCCCCAGTG GTGACCTATG TGTGTGATCT
TGAAGGGAAC AGAGTCAAGG GTCCAGAGAA GGAGGAGAAG TTGAGACAGG 400
CGGTCAAGCA GGTGCTGAAG TGTGATGTGA CTCAGAGCCA GCCACTGGGG
                          C
GCCGTCCCCT TACCCCCGGC TGACTGCGTG CTCAGCACAC TGTGTCTGGA 500
TGCCGCCTGC CCAGACCTCC CCACCTACTG CAGGGCGCTC AGGAACCTCG
GCAGCCTACT GAAGCCAGGG GGCTTCCTGG TGATCATGGA TGCGCTCAAG 600
AGCAGCTACT ACATGATTGG TGAGCAGAAG TTCTCCAGCC TCCCCCTGGG
CCGGGAGGCA GTAGAGGCTG CTGTGAAAGA GGCTGGCTAC ACAATCGAAT 700
GGTTTGAGGT GATCTCGCAA AGTTATTCTT CCACCATGGC CAACAACGAA
GGACTTTTCT CCCTGGTGGC GAGGAAGCTG AGCAGACCCC TGTGA      795
```

FIGURE 2

AMINO ACID SEQUENCE OF THE NNMT PROTEIN

```
MESGFTSKDT YLSHFNPRDY LEKYYKFGSR HSAESQILKH LLKNLFKIFC
LDGVKGDLLI DIGSGPTIYQ LLSACESFKE IVVTDYSDQN LQELEKWLKK 100
EPEAFDWSPV VTYVCDLEGN RVKGPEKEEK LRQAVKQVLK CDVTQSQPLG
AVPLPPADCV LSTLCLDAAC PDLPTYCRAL RNLGSLLKPG GFLVIMDALK 200
SSYYMIGEQK FSSLPLGREA VEAAVKEAGY TIEWFEVISQ SYSSTMANNE
GLFSLVARKL SRPL                                       264
```

FIGURE 3

SEQUENCE LISTING

<110> Genaissance Pharmaceuticals, Inc.
Chew, Anne
Gilson, Christopher
Kazemi, Amir
Koshy, Beena

<120> HAPLOTYPES OF THE NNMT GENE

<130> MWH-0179PCT

<140> TBA
<141> 2002-05-07

<150> 30/289,335
<151> 2001-05-07

<160> 21

<170> PatentIn version 3.1

<210> 1
<211> 3385
<212> DNA
<213> Homo sapiens

<220>
<221> allele
<222> (394)..(394)
<223> PS1: polymorphic base adenine or thymine

<220>
<221> allele
<222> (928)..(928)
<223> PS2: polymorphic base thymine or cytosine

<220>
<221> misc_feature
<222> (2433)..(2532)
<223> N's represent unknown nucleotides

<220>
<221> allele
<222> (2696)..(2696)
<223> PS3: polymorphic base thymine or cytosine

<400> 1
cagacactgg gtcatggcag tggtcggtga agctgcagtt gcctagggca gggatggaga 60
gagagtctgg gcatgaggag agggtctcgg gatgtttggc tggactagat tttacagaaa 120
gccttatcca ggcttttaaa attactcttt ccagacttca tctgagactc cttcttcagc 180
caacattcct tagccctgaa tacatttcct atcctcatct ttcccttctt ttttttcctt 240
tcttttacat gtttaaattt aaaccattct tcgtgacccc ttttcttggg agattcatgg 300
caagaacgag aagaatgatg gtgcttgtta ggggatgtcc tgtctctctg aactttgggg 360
tcctatgcaa taaataattt tcctgacgag ctcwagtgct ccctctggtc tacaatccct 420
ggcggctggc cttcatccct tgggcaagca ttgcatacag ctcatggccc tccctctacc 480
ataccctcca cccccgttcg cctaagctcc cttctccggg aatttcatca tttcctagaa 540
cagccagaac atttgtggtc tatttctctg ttagtgttta accaaccatc tgttctaaaa 600
gaagggctga actgatggaa ggaatgctgt tagcctgaga ctcaggaaga caacttctgc 660
agggtcactc cctggcttct ggaggaaaga gaaggagggc agtgctccag tggtacagaa 720
gtgagacata atggaatcag gcttcacctc caaggacacc tatctaagcc attttaaccc 780
tcgggattac ctagaaaaat attacaagtt tggttctagg cactctgcag aaagccagat 840
tcttaagcac cttctgaaaa atcttttcaa gatattctgc ctaggtaagt ctgttgtctg 900
catgtctccc cactaatgtg agtcatayag atggagtctc agggcacgac tgggttttgt 960
gtctctcgtt gttgcttcac agccctttg gcatcaccca tttatttaac taggataaaa 1020
acgaatattg gtatagcgat tccacagttt acaaagtgct tttgtatcca ctgtctcact 1080
tgatcaagca aaaggaaacc agaggaccgg agtgctgtcc tgagtctacc ttgatttgct 1140
aggcgacttg agggagactt ttagcctcaa agggccactt aagtggaaat tctaaaacag 1200
tacctattct gatcctaact caagggaatg ctgtgaatat gcatgagata aagacctccc 1260

aatatatgaa gaactgggtg attttgggag aaagacatta tatactcaat ttctttttta 1320
attaactttc cttgaaagta ttgcttaata gtttttacat tctccatgta acagactttc 1380
tggatctggt cttcagtctg tacaccagat gtagatcttt tttaccttct cctagacctt 1440
aaaattcctg gcaacatgcc tccaccctgg attggggaat aaaaaatgaa aagtttttt 1500
ttttctttt tgactttaaa ttttattaaa gtttgaggtt tttcaaactg atgtgcttta 1560
tttaaaatcc aagtgagaca tttttagtct ttttgatatt tatattttct ttgtcactat 1620
gatgtaaatt acagggattt gggggaaaaa tgggattttt ttttttttt tggagatata 1680
gatctcactc tgtttcctag gctggatgga gtgcagggat gtgatcacag ctcatcatag 1740
tctcgaactc ctggactcaa gggatccttc tgcctcagcc tctccaataa ctaggtcttc 1800
aggcacacgc caccatgcct agctaatttt aaaatttttt tgtagagatg aggtctcact 1860
atgttgtcca ggctggtctc atcctccagg cctcaagtga tcctcctgcc ttggccttcc 1920
agagtattgg gactgtaggc atgagccact gtgcctggcc cagaaaagat gttttaaaaa 1980
aacattttga gggaaaagtt gtgaacagta gtggtctgtc tttgaggatc gccagcacag 2040
tcccagggaa gacaatgtaa atttgactct gcccactgcc atgagatgcc tgatctctcc 2100
tctttgttcc tcccactaat ccagacggtg tgaagggaga cctgctgatt gacatcggct 2160
ctggccccac tatctatcag ctcctctctg cttgtgaatc ctttaaggag atcgtcgtca 2220
ctgactactc agaccagaac ctgcaggagc tggagaagtg gctgaagaaa gagccagagg 2280
cctttgactg gtccccagtg gtgacctatg tgtgtgatct tgaagggaac aggtagagaa 2340
actggtgtct acttcttggc ttttgaaggt acctgagtga tggttggcaa aagcaacaga 2400
cagataggga ccaaagagaa atccaaatgg agnnnnnnnn nnnnnnnnnn nnnnnnnnnn 2460
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 2520
nnnnnnnnnn nncacatctg ggacaatacg gccatttttа gccttgaccc aagagatctg 2580
ggttccccat gactggagtg gaaaacaatg tctgtgggtt tgtgtttttc agagtcaagg 2640
gtccagagaa ggaggagaag ttgagacagg cggtcaagca ggtgctgaag tgtgaygtga 2700
ctcagagcca gccactgggg gccgtcccct tacccccggc tgactgcgtg ctcagcacac 2760
tgtgtctgga tgccgcctgc ccagacctcc ccacctactg cagggcgctc aggaacctcg 2820
gcagcctact gaagccaggg ggcttcctgg tgatcatgga tgcgctcaag agcagctact 2880
acatgattgg tgagcagaag ttctccagcc tccccctggg ccgggaggca gtagaggctg 2940
ctgtgaaaga ggctggctac acaatcgaat ggtttgaggt gatctcgcaa agttattctt 3000
ccaccatggc caacaacgaa ggacttttct ccctggtggc gaggaagctg agcagacccc 3060
tgtgatgcct gtgacctcaa ttaaagcaat tcctttgacc tgtccagttg actttagtcc 3120
ttgtttctaa ctgccaagtc atgtgctgag tagaggctca gtggttgggg cccaatggtt 3180
catctaggac gggactagag aggtcagtct acaagcaatc cattgaccac ttacttggtg 3240
ctgcacacaa atgttggtgc tatgggaccc aaagatgagc aattagtatt ccagtcttca 3300
ttgcctgtgc ttacaaaaga agacctcact tccctaaaca tctagttatg gaggttcaag 3360
cccgtacctg cctacagaga agtgt 3385

<210> 2
<211> 795
<212> DNA
<213> Homo sapiens

<400> 2
atggaatcag gcttcacctc caaggacacc tatctaagcc attttaaccc tcgggattac 60
ctagaaaaat attacaagtt tggttctagg cactctgcag aaagccagat tcttaagcac 120
cttctgaaaa atcttttcaa gatattctgc ctagacggtg tgaagggaga cctgctgatt 180
gacatcggct ctggccccac tatctatcag ctcctctctg cttgtgaatc ctttaaggag 240
atcgtcgtca ctgactactc agaccagaac ctgcaggagc tggagaagtg gctgaagaaa 300
gagccagagg cctttgactg gtccccagtg gtgacctatg tgtgtgatct tgaagggaac 360
agagtcaagg gtccagagaa ggaggagaag ttgagacagg cggtcaagca ggtgctgaag 420
tgtgatgtga ctcagagcca gccactgggg gccgtcccct tacccccggc tgactgcgtg 480
ctcagcacac tgtgtctgga tgccgcctgc ccagacctcc ccacctactg cagggcgctc 540
aggaacctcg gcagcctact gaagccaggg ggcttcctgg tgatcatgga tgcgctcaag 600
agcagctact acatgattgg tgagcagaag ttctccagcc tccccctggg ccgggaggca 660
gtagaggctg ctgtgaaaga ggctggctac acaatcgaat ggtttgaggt gatctcgcaa 720
agttattctt ccaccatggc caacaacgaa ggacttttct ccctggtggc gaggaagctg 780
agcagacccc tgtga 795

<210> 3
<211> 264
<212> PRT
<213> Homo sapiens

<400> 3
Met Glu Ser Gly Phe Thr Ser Lys Asp Thr Tyr Leu Ser His Phe Asn
1 5 10 15

Pro Arg Asp Tyr Leu Glu Lys Tyr Tyr Lys Phe Gly Ser Arg His Ser
20 25 30

```
Ala Glu Ser Gln Ile Leu Lys His Leu Leu Lys Asn Leu Phe Lys Ile
        35                  40                  45

Phe Cys Leu Asp Gly Val Lys Gly Asp Leu Leu Ile Asp Ile Gly Ser
    50                  55                  60

Gly Pro Thr Ile Tyr Gln Leu Leu Ser Ala Cys Glu Ser Phe Lys Glu
65                  70                  75                  80

Ile Val Val Thr Asp Tyr Ser Asp Gln Asn Leu Gln Glu Leu Glu Lys
                85                  90                  95

Trp Leu Lys Lys Glu Pro Glu Ala Phe Asp Trp Ser Pro Val Val Thr
            100                 105                 110

Tyr Val Cys Asp Leu Glu Gly Asn Arg Val Lys Gly Pro Glu Lys Glu
        115                 120                 125

Glu Lys Leu Arg Gln Ala Val Lys Gln Val Leu Lys Cys Asp Val Thr
    130                 135                 140

Gln Ser Gln Pro Leu Gly Ala Val Pro Leu Pro Pro Ala Asp Cys Val
145                 150                 155                 160

Leu Ser Thr Leu Cys Leu Asp Ala Ala Cys Pro Asp Leu Pro Thr Tyr
                165                 170                 175

Cys Arg Ala Leu Arg Asn Leu Gly Ser Leu Leu Lys Pro Gly Gly Phe
            180                 185                 190

Leu Val Ile Met Asp Ala Leu Lys Ser Ser Tyr Tyr Met Ile Gly Glu
        195                 200                 205

Gln Lys Phe Ser Ser Leu Pro Leu Gly Arg Glu Ala Val Glu Ala Ala
    210                 215                 220

Val Lys Glu Ala Gly Tyr Thr Ile Glu Trp Phe Glu Val Ile Ser Gln
225                 230                 235                 240

Ser Tyr Ser Ser Thr Met Ala Asn Asn Glu Gly Leu Phe Ser Leu Val
                245                 250                 255

Ala Arg Lys Leu Ser Arg Pro Leu
            260
```

<210> 4
<211> 15
<212> DNA
<213> Homo sapiens

<400> 4
cgagctcwag tgctc 15

<210> 5
<211> 15
<212> DNA
<213> Homo sapiens

<400> 5
agtcatayag atgga 15

<210> 6
<211> 15
<212> DNA
<213> Homo sapiens

<400> 6
agtgtgaygt gactc 15

<210> 7
<211> 15
<212> DNA
<213> Homo sapiens

<400> 7
tcctgacgag ctcwa 15

<210> 8
<211> 15
<212> DNA
<213> Homo sapiens

<400> 8
cagagggagc actwg 15

<210> 9
<211> 15
<212> DNA
<213> Homo sapiens

<400> 9
aatgtgagtc ataya 15

<210> 10
<211> 15
<212> DNA
<213> Homo sapiens

<400> 10
tgagactcca tctrt 15

<210> 11
<211> 15
<212> DNA
<213> Homo sapiens

<400> 11
tgctgaagtg tgayg 15

<210> 12
<211> 15
<212> DNA
<213> Homo sapiens

<400> 12
ggctctgagt cacrt 15

<210> 13
<211> 10
<212> DNA
<213> Homo sapiens

<400> 13
tgacgagctc 10

<210> 14
<211> 10
<212> DNA
<213> Homo sapiens

<400> 14
agggagcact 10

<210> 15
<211> 10
<212> DNA
<213> Homo sapiens

<400> 15
gtgagtcata 10

<210> 16
<211> 10
<212> DNA
<213> Homo sapiens

<400> 16
gactccatct 10

<210> 17
<211> 10
<212> DNA
<213> Homo sapiens

<400> 17
tgaagtgtga 10

<210> 18
<211> 10
<212> DNA
<213> Homo sapiens

<400> 18
tctgagtcac 10

<210> 19
<211> 18
<212> DNA
<213> Homo sapiens

<400> 19
tgtaaaacga cggccagt 18

<210> 20
<211> 19
<212> DNA
<213> Homo sapiens

<400> 20
aggaaacagc tatgaccat 19

<210> 21
<211> 360
<212> DNA
<213> Homo sapiens

<220>
<221> allele
<222> (30)..(30)
<223> PS1: polymorphic base adenine or thymine

<220>
<221> misc_feature
<222> (61)..(120)
<223> N's represent nucleotides between PS1 and PS2

<220>
<221> allele
<222> (150)..(150)
<223> PS2: polymorphic base thymine or cytosine

<220>
<221> misc_feature
<222> (181)..(240)
<223> N's represent nucleotides between PS2 and PS3

<220>
<221> allele
<222> (270)..(270)
<223> PS3: polymorphic base thymine or cytosine

<220>
<221> misc_feature
<222> (301)..(360)
<223> N's represent nucleotides 3' of PS3

<400> 21
atgcaataaa taattttcct gacgagctcw agtgctccct ctggtctaca atccctggcg 60
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 120
tgcatgtctc cccactaatg tgagtcatay agatggagtc tcagggcacg actgggtttt 180
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 240
caggcggtca agcaggtgct gaagtgtgay gtgactcaga gccagccact gggggccgtc 300
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 360